



OPEN

Ultrasound and SWE-based transfer learning for predicting fibrotic NASH

Fei Xia^{1,4}, Kun Wang^{1,4}, Yuhe Wang^{1,2}, Chaoxue Zhang³ & Junli Wang^{1,2✉}

The aim of this study was to develop a combined deep-learning model utilizing liver ultrasound, liver elastography images, and clinical features to predict and diagnose fibrotic non-alcoholic steatohepatitis (NASH). A rat model of liver steatosis and fibrosis was established through a high-fat diet and subcutaneous CCl₄ injections. Two-dimensional ultrasound and shear wave elastography (SWE) images were acquired. Three deep learning models, based on the ResNet-18 architecture, were designed: (1) a pure image model using only liver ultrasound, (2) a pure image model using only liver elastography, and (3) a combined model incorporating liver ultrasound, liver elastography images, and clinical features. The performance of these models was evaluated using three-fold cross-validation, receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and calibration curves. The combined deep learning model demonstrated the highest area under the curve (AUC) of 0.879. DCA revealed that the multimodal model provided superior net benefits across most threshold probability ranges for predicting and diagnosing fibrotic NASH. The combined deep learning model based on the ResNet-18 architecture exhibits promising performance in predicting and diagnosing fibrotic NASH.

Keywords Liver fibrosis, Non-alcoholic steatohepatitis (NASH), Deep learning, Real-time shear wave elastography

Non-alcoholic fatty liver disease (NAFLD), driven by the global prevalence of obesity and type 2 diabetes mellitus (T2DM), has become the most common cause of chronic liver disease worldwide¹. The primary characteristic of NAFLD is fat accumulation in the liver parenchyma, which can progress to a more severe inflammatory stage characterized by hepatocellular injury with or without fibrosis, known as non-alcoholic steatohepatitis (NASH)². NASH is a risk factor for liver fibrosis, which indicates a poor prognosis. However, not all NASH patients progress to fibrosis³. Previous studies have demonstrated that patients with fibrosis-related NASH are at significantly increased risk of liver-related complications and progression to cirrhosis^{4,5}. Additionally, liver fibrosis has been identified as the strongest predictor of disease-specific progression and mortality^{6,7}. Consistent with previous research, we classify NASH patients with a fibrosis stage ≥ 2 as fibrotic NASH^{8,9}. These patients face the highest risk of disease progression, benefit the most from pharmacological treatment, and are the population most in need of therapeutic intervention. The timely diagnosis of fibrotic NASH patients is of great significance for the clinical management of liver disease.

Currently, non-invasive clinical methods for assessing and evaluating NAFLD, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, are ineffective in accurately detecting fibrotic NASH patients¹⁰. In recent years, artificial intelligence (AI) has emerged as a powerful tool in medical diagnosis, particularly in medical image analysis. Deep learning, a subset of AI based on deep neural networks (DNNs) with multiple hidden layers, is at the forefront of this field¹¹. Neural networks, inspired by the structure and function of biological neurons, possess non-linear, adaptive, and parallel processing capabilities, making them well-suited for multi-parameter diagnostic models. These models have demonstrated promising results in predicting, classifying, and assessing liver fibrosis^{12,13}. Deep learning is a rapidly evolving technology with tremendous potential in various areas of medical image analysis, including disease classification, segmentation, detection, and image registration. Existing studies have shown that multimodal deep learning models can enhance model performance. For example, Sato et al. integrated ultrasound images with clinical information

¹Department of Ultrasound, The Second People's Hospital, WuHu Hospital, East China Normal University, WuHu), No.259 Jiuhuashan Road, Jinghu District, Wuhu 241001, Anhui, China. ²School of Graduate Studies, Bengbu Medical College, No. 2600 Donghai Avenue, Longzihu District, Bengbu 233030, Anhui Province, China. ³Department of Ultrasound, The First Affiliated Hospital of Anhui Medical University, No.21 Jixi Road, Shushan District, Hefei 230022, Anhui, China. ⁴Fei Xia and Kun Wang contributed equally to this work. ✉email: wjl980134@163.com

such as patient age and blood biochemical indicators, resulting in an improvement in diagnostic accuracy from 68.52% to 96.30% (AUC = 0.994)¹⁴. In addition, deep learning models can expand their application scope from diagnosis to treatment decision-making. For instance, Chi Zhang et al. developed an acute pancreatitis model that innovatively combines lesion segmentation with severity assessment to assist in formulating personalized treatment plans¹⁵. Moreover, deep learning models can be combined with Grad-CAM to generate heatmaps, visually illustrating diagnostic reasoning. For example, the model developed by Hong et al. used Grad-CAM to highlight key regions associated with visual impairment diseases¹⁶. These advancements further validate the effectiveness and cutting-edge nature of deep learning-based diagnostic models.

Currently, deep learning models employed for ultrasound diagnosis primarily rely on two-dimensional cross-sectional images for predictive tasks. However, the information derived from two-dimensional cross-sections is inherently limited and may not fully capture the complexity of lesions. Real-Time Shear Wave Elastography (SWE) is an ultrasound-based technique that quantifies tissue stiffness by measuring tissue elasticity, thereby enabling the assessment of tissue characteristics¹⁷. Previous research has substantiated the efficacy of SWE in detecting fibrosis^{18,19}. This study aimed to develop a combined deep-learning model that leverages liver ultrasound, liver elastography images, and clinical features to assess the risk of fibrotic NASH. The performance of this combined model was compared with deep learning models trained exclusively on 2D ultrasound images and models trained solely on elastography images.

Materials and methods

Animal model

This study was approved by the Animal Care and Use Committee of our hospital and complied with the national standards for animal care and use in China. All procedures were conducted in accordance with the ARRIVE guidelines, and all methods were performed in accordance with the relevant guidelines and regulations. Ethical approval number: 2024-KY-112. A total of ninety male Sprague-Dawley (SD) rats (initial weight: 140–170 g) were used. The rats used in this study were obtained from the Pengyue Experimental Animal Breeding Co., Ltd. in Jinan, China. After a three-day acclimatization period, the rats were randomly divided into two groups: a normal control group ($n=30$) fed a standard diet and a high-fat diet group ($n=60$) fed a high-fat diet. The dietary energy composition was as follows: 20% protein, 20% carbohydrates, and 60% fat. The high-fat diet group received a CCl₄ oil mixture (1:4) from week 8 to week 14 (0.2 mL/kg, twice weekly, intraperitoneally). In the control group, 10 rats were randomly selected at weeks 4, 10, and 14, while in the high-fat diet group, 10 rats were randomly selected at weeks 4, 6, 8, 10, 12, and 14. After intraperitoneal injection of pentobarbital sodium for anesthesia, imaging examinations were performed, and blood samples were collected from the abdominal aorta. Subsequently, under anesthesia, cervical dislocation was performed to euthanize the rats, and liver tissue was collected.

Multimodal ultrasound image acquisition

Before the acquisition of ultrasound images, all rats were fasted for 24 h. Anesthesia was induced by intraperitoneal injection of sodium pentobarbital (3% saline solution, 40 mg/kg). Following successful anesthesia, the rats' abdomens were shaved, and the animals were secured on the operating table for ultrasound scanning. Ultrasound examinations were performed using an ACUSON Sequoia real-time shear wave elastography ultrasound diagnostic system (Siemens, USA) equipped with a conventional linear array 10L4 probe (4–10 MHz). The examination commenced with routine B-mode ultrasound scanning. After the standard B-mode scan, 2D images of the liver's median lobe, left lateral lobe, and left medial lobe were consistently acquired and stored. Subsequently, the modality was switched to 2D-shear wave elastography (SWE) mode to obtain elastography data for the same liver lobes. These ultrasound images, encompassing both grayscale ultrasound and SWE results, were archived in the Digital Imaging and Communications in Medicine (DICOM) format for subsequent radiomics analysis.

Serum and tissue pathological analysis

After completing the ultrasound scans, the rats' body weight was measured. Subsequently, blood was collected via puncture of the abdominal aorta. The rats were then euthanized by cervical dislocation under anesthesia. The liver was rapidly dissected, washed with physiological saline, and the median lobe was fixed in 10% neutral formalin for subsequent analysis. The collected serum samples were appropriately labeled and sent to the hospital's laboratory for analysis. The tests included measurement of alanine aminotransferase (ALT), aspartate aminotransferase (AST), AST/ALT ratio, gamma-glutamyl transferase (GGT), total cholesterol (CHOL), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C).

The liver tissue was fixed in 10% neutral formalin for 24 h, followed by conventional tissue processing and hematoxylin-eosin (HE) staining. Pathological diagnosis of NAFLD was based on the Steatosis, Activity, and Fibrosis (SAF) scoring system. NASH is characterized by hepatocellular fat accumulation, ballooning degeneration, and portal inflammation, each scored on a scale of 0–3. A total SAF score ≥ 3 indicates NASH. Significant fibrosis is defined as a fibrosis stage $\geq F2$ (F0: no fibrosis; F1: peri-sinusoidal or peri-portal fibrosis; F2: peri-sinusoidal and portal/peri-portal fibrosis; F3: bridging fibrosis; F4: cirrhosis). Fibrotic NASH is diagnosed when both NASH and significant fibrosis are present.

Deep learning model development

This study designed three deep learning models: a pure image model utilizing solely liver ultrasound images, a pure image model employing exclusively liver elastography images, and a combined model integrating both liver ultrasound and liver elastography images, along with pertinent clinical features.

The deep learning model employed in this study is based on the ResNet-18 architecture, which leverages residual modules to effectively mitigate degradation and vanishing gradient problems during the training of deep networks. We constructed three separate models: a unimodal 2D ultrasound model, a unimodal SWE model, and a multimodal fusion model. For the fusion model, features were first extracted independently from 2D ultrasound and SWE images using separate network branches, and then fused at the image level via concatenation along the channel dimension. Additionally, clinical features were mapped into a 100-dimensional vector through a fully connected layer and subsequently concatenated with the 512-dimensional image features output from the ResNet branches. These combined features were further integrated using a self-attention mechanism to enhance the model's ability to capture heterogeneous information, and finally passed through a fully connected classifier to generate predictions.

For model training, all networks were initialized with pre-trained ResNet-18 parameters. The Adam optimizer was employed, with an initial learning rate set to 0.0001 and a batch size of 8. During training, we monitored the loss on the validation set and applied early stopping if no loss reduction was observed within 30 epochs, in order to prevent overfitting. The training process was implemented using the PyTorch framework on an NVIDIA GeForce RTX 4060 Ti GPU. To enhance the generalization capability of the model, a variety of commonly used data augmentation techniques were applied to all input ultrasound and SWE images. Specifically, random horizontal flipping was applied with a probability of 50%, meaning that half of the training samples were flipped left-to-right to increase data diversity. In addition, images were randomly cropped with a scaling factor ranging from 0.8 to 1.0 and then resized back to the standard input size required by the model, simulating variations in the field of view during image acquisition. Brightness, contrast, saturation, and hue were also randomly adjusted with predefined probabilities: brightness and contrast were varied within $\pm 20\%$, saturation within $\pm 10\%$, and hue within ± 5 degrees. All augmentations were performed online during training, allowing the model to encounter a wider range of image variations in each epoch. This reduces reliance on specific acquisition conditions and improves the model's robustness to real-world data. These augmentation strategies are consistent with existing literature and standard practices in medical image processing, ensuring the validity of the samples while effectively mitigating the risk of overfitting due to limited dataset size.

Univariate analysis was conducted to identify indicators associated with fibrosis, followed by multivariate logistic regression analysis to determine independent predictors. To predict fibrotic non-alcoholic steatohepatitis based on ultrasound images alone or in combination with clinical factors, we employed various deep-learning architectures. For the pure image model, a ResNet18 network with a fully connected layer was utilized to extract high-dimensional features from the imaging data and directly predict the probability of fibrotic non-alcoholic steatohepatitis. For the combined model ResNet-based Hybrid Network(ResHy-Net), as shown in Fig. 1, a Convolutional Neural Network (CNN) network was employed as a powerful feature extractor, with the final layer and classification layer removed to obtain necessary feature maps. Ultrasound features from different modalities were extracted using the CNN network and concatenated along the channel dimension. Clinical features were transformed into a 100-dimensional vector through a fully connected layer and concatenated with the 512-dimensional ultrasound image features. This combined feature vector was then input into a multi-head self-attention module (Fig. 2), where features were segmented into fixed-size blocks and subjected to multi-head self-attention calculations to incorporate positional information. The processed blocks were subsequently classified using a classifier, with the fused information decoded through a multi-layer perceptron to predict the fibrosis status of the rat liver. To address the issue of overfitting due to data imbalance, a weighted oversampling method and binary cross-entropy loss function were employed.

Considering the imbalance between fibrotic NASH and non-fibrotic NASH samples in our study, we employed a combined strategy of weighted oversampling and loss function reweighting to effectively mitigate model bias caused by class imbalance. Specifically, during the oversampling phase, we adopted the WeightedRandomSampler

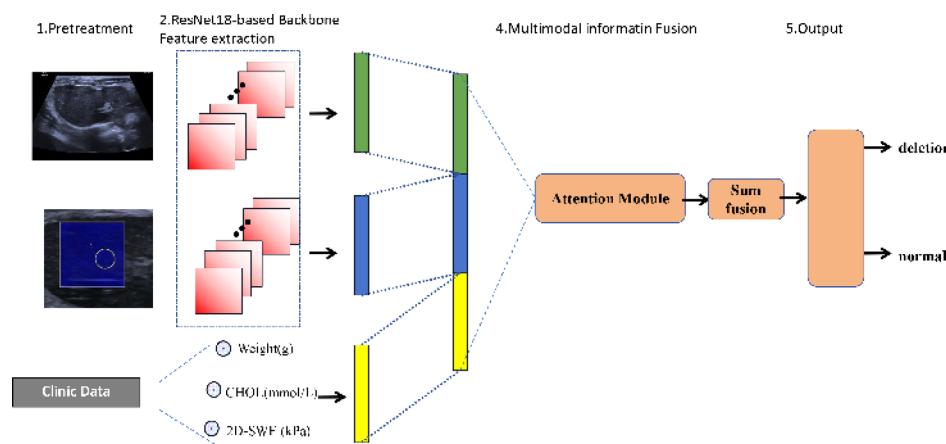


Fig. 1. Overview of the ResHy Network. Using CNNs, features from 2D ultrasound and SWE images were extracted separately, and the feature maps were concatenated. A spatial attention module was then employed to fuse the features from different sequences along with clinical characteristics. Finally, the liver's status was determined by a classifier, specifically a multi-layer perceptron.

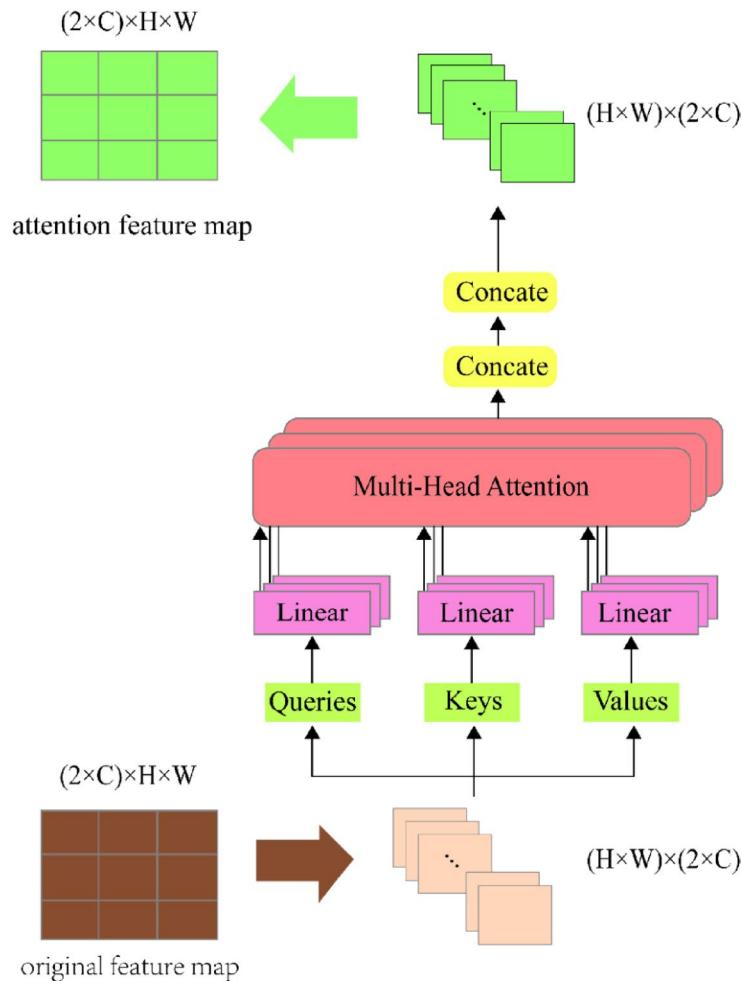


Fig. 2. The details of spatial attention module. Split the feature map into fixed-size patches and perform multi-head self-attention calculation on these different patches to obtain a feature map with position information.

implemented in PyTorch, with sampling weights assigned as the inverse of the class frequencies in the training set. In other words, since fibrotic NASH samples are underrepresented, each instance in this class was given a higher probability of being selected during training. This approach ensures a more balanced distribution of modality data within each training epoch. Unlike naive oversampling methods that may repeatedly duplicate rare samples, our strategy achieves distributional rebalancing while preserving the full diversity of the dataset and avoiding overfitting to individual instances.

At the loss function level, to further enhance the model's discriminative ability for the minority class, we incorporated a weighting scheme into the conventional binary cross-entropy (BCE) loss, resulting in a weighted binary cross-entropy loss. Specifically, the loss assigns weights inversely proportional to the sample counts of each class: fibrotic NASH (the minority class) samples are given higher penalty weights (e.g., the positive class weight is set as the ratio of negative samples to total samples, and vice versa for the negative class). This ensures that during training error optimization, the model does not become overly biased toward the majority class.

This combined strategy of weighted sampling and weighted BCE loss ensures that the model not only increases the training frequency of minority class samples at the data sampling level but also heightens the sensitivity to misclassification of minority classes at the loss penalty level. It is a well-established, mainstream, and effective approach for addressing class imbalance in medical datasets.

The weighted binary cross-entropy loss (Weighted BCE) we employed is defined as follows:

$$L = -\frac{1}{N} \sum_{i=1}^N [w_1 \cdot y_i \cdot \log(p_i) + w_0 \cdot (1 - y_i) \cdot \log(1 - p_i)]$$

where N is the total number of samples, y_i is the true label of the i -th sample ($y_i = 1$ for fibrotic NASH, and $y_i = 0$ for non-fibrotic NASH), p_i is the predicted probability from the model, and w_1 and w_0 are the loss weights for the fibrotic NASH (positive class) and non-fibrotic NASH (negative class), respectively. These weights are set as the inverse of the class sample counts:

$$w_1 = \frac{N}{2N_1}, w_0 = \frac{N}{2N_0}$$

Where N1 and N0 represent the number of positive and negative samples, respectively. This weighting scheme ensures that during optimization, greater emphasis is placed on the minority class samples.

At the sampling level, we adopted the WeightedRandomSampler based on the above weights w_1 and w_0 in order to increase the occurrence probability of minority class samples during training.

Implementation of CNN and cross-validation

To evaluate the performance of ResNet-based Hybrid Network(ResHy-Net), an independent network based on CNN ResNet18, the dataset was divided into training and validation sets. In our study, the train-validation split was performed at the animal level rather than the image level. Specifically, we ensured that all images derived from the same rat were assigned exclusively to either the training set or the validation set, never to both. This approach was adopted to strictly prevent data leakage and to ensure that the evaluation of model performance reflects its generalization to unseen subjects rather than repeated sampling of the same individual. In addition, we employed three-fold cross-validation to evaluate the performance of the model.

Initially, three rounds of training were conducted using individual ultrasound images. Subsequently, three rounds of training were performed with multimodal ultrasound images. In each round, one group served as the validation set, while the remaining data was used for training. A separate model was generated from each training session. The overall performance was evaluated by averaging the validation results from the three models. This approach mitigated the impact of uneven data distribution and provided a more robust assessment of the model's performance. Image augmentation techniques, including flipping, cropping, and adjustments to brightness, contrast, saturation, and hue, were applied to the training images. The network was implemented using PyTorch with the Adam (Adaptive Moment Estimation) optimizer. The initial learning rate was set to 10^{-4} , and the batch size was 8. An early stopping strategy was implemented, terminating training if the validation loss did not decrease for 30 epochs. The ResNet parameters were initialized using pre-trained weights from the ImageNet dataset. Network training was conducted on an NVIDIA GeForce RTX 4060 Ti GPU.

Statistical analysis

To evaluate the overall model accuracy, we used the output probabilities from the softmax layer of the model and applied a threshold of 0.5 to determine the fibrosis status. An average softmax score greater than 0.5 indicated a prediction of fibrotic status, while an average softmax score less than 0.5 indicated a prediction of non-fibrotic status. Although 0.5 is the standard threshold for neural network classification, this threshold can be adjusted between 0 and 1 to modify the network's sensitivity and specificity. Using this method, algorithm performance was evaluated by calculating the area under the curve (AUC) across different threshold ranges. To assess the algorithm's generalizability, a three-fold cross-validation approach was employed. In each iteration, one-third of the data was used for validation, while the remaining two-thirds were used for training. The reported results are from the validation sets after three-fold cross-validation. Finally, the model's performance was evaluated using ROC curves. Decision curve analysis (DCA) was employed to compare the clinical value of various predictive models, and calibration curves were used to assess model fit. DCA calculations were primarily performed using the R software packages "rms" and "rmda". A *p*-value of less than 0.05 was considered statistically significant for all two-sided tests.

Results

The clinical characteristics of rats

Three of the 90 rats died unexpectedly, leaving 87 rats for further analysis. Table 1 presents the basic clinical information of the remaining rats. Among these, 23 rats were diagnosed with fibrotic NASH, and 64 rats were diagnosed with non-fibrotic NASH. The mean weight of rats with non-fibrotic NASH was approximately 479.4 ± 90.2 g, while the average weight of rats with fibrotic NASH was approximately 583.3 ± 62.5 g. Significant differences ($p < 0.05$) were observed between the non-fibrotic NASH and fibrotic NASH rat groups in terms of weight, ALT, AST, GGT, CHOL, LDL-C, and SWE. To identify independent clinical predictors, multivariate

Characteristics	Non-Fibrotic -NASH	Fibrotic -NASH	P
Weight(g), mean (SD)	479.4 ± 90.2	583.3 ± 62.5	<0.001
ALT(U/L), median [Q1,Q3]	37.0 [30.0,66.5]	113.0 [62.5,302.0]	<0.001
AST(U/L), median [Q1,Q3]	86.0 [68.8,200.2]	440.0 [186.0,580.5]	<0.001
AST/ALT, median [Q1,Q3]	2.6 [2.1,3.6]	2.8 [2.2,3.5]	0.616
GGT(U/L), median [Q1,Q3]	1.0 [1.0,2.0]	2.0 [1.5,11.0]	0.023
CHOL(mmol/L), mean (SD)	1.9 (0.4)	1.7 (0.5)	0.030
TG(mmol/L), mean (SD)	0.7 (0.3)	0.8 (0.3)	0.613
HDL-C(mmol/L), mean (SD)	1.0 (0.4)	0.8 (0.2)	0.063
LDL-C(mmol/L), mean (SD)	0.3 (0.1)	0.4 (0.2)	0.004
2D-SWE (kPa), median [Q1,Q3]	4.3 [3.8,4.8]	6.2 [5.1,6.8]	<0.001

Table 1. Clinical characteristics.

logistic regression analysis was performed on the relevant clinical factors. As shown in Table 2, weight, CHOL, and SWE ($p < 0.05$) were identified as independent clinical predictive factors.

Interpretability of deep learning radiomics models

Explainable artificial intelligence can enhance model transparency and trust, providing better insights into the underlying decision-making processes. This is crucial for ensuring fairness, avoiding bias, and maintaining safety in critical fields such as healthcare and finance²⁰. To better understand the decision-making basis of the ResNet-based Hybrid Network(ResHy-Net) model, this study employed Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the response areas of the deep learning model when assessing fibrotic non-alcoholic steatohepatitis. Significance maps were obtained by applying Grad-CAM to the last convolutional layer of the CNN. Additionally, we calculated the relative weights of each clinical risk factor in the combined model (Fig. 3).

Comparison of different deep learning models

The average cross-validation test accuracy (ACC) of the ResNet-based Hybrid Network(ResHy-Net) model trained on individual 2D ultrasound images was 0.776 ± 0.0139 , The average cross-validation test area under the curve value was 0.736 ± 0.0526 , The average cross-validation test ACC of the ResNet-based Hybrid Network(ResHy-Net) model trained on individual SWE ultrasound images was 0.776 ± 0.0139 , and the average cross-validation test AUC value was 0.776 ± 0.066 . The combined model, integrating 2D ultrasound, SWE images, and independent clinical predictors, achieved an ACC of 0.880 ± 0.0372 , and an average cross-validation AUC of 0.879 ± 0.0316 . A comparison of the accuracy and AUC scores of the three models is presented in Table 3. The performance metrics of all three models, including precision, recall, sensitivity, and specificity, are presented in Supplementary Tables S1–S3. We also included models based solely on clinical features, elasticity values (SWE), and a 2D + SWE deep learning model. All of these models demonstrated inferior performance compared to the combined model. The performance metrics of these models are presented in Table 4.

The ROC curves for each cross-validation of the three models are presented in Fig. 4. The combined model exhibited exceptionally high sensitivity and specificity. The clinical utility of the models was directly evaluated using DCA. The DCA curves of the standalone 2D and standalone SWE deep learning models are shown in Supplementary Figures S1–S2. The DCA curve for the combined deep learning model (Fig. 5) demonstrates that, across most threshold probability ranges, this model offers a greater net benefit in predicting and diagnosing fibrotic NASH. Additionally, the calibration curve was utilized to assess the prediction performance and accuracy of the multimodal model. In the training set, the average p-value of the Hosmer-Lemeshow (H-L) test for this model was 0.24 (0.26, 0.23, 0.23). In the validation set, the p-values of the H-L test for the combined model were 0.30 (0.06, 0.31, 0.54). These results indicate that the multimodal model is highly suitable for both the training set and the validation set (Fig. 6).

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	p	OR (95% CI)	p
Weight	1.014 (1.007–1.021)	<0.001	1.016 (1.002–1.030)	0.022
ALT	1.012 (1.006–1.018)	<0.001	0.999 (0.982–1.018)	0.987
AST	1.003 (1.001–1.005)	<0.001	1.001 (0.992–1.010)	0.867
AST/ALT	0.941 (0.743–1.192)	0.616	NA	NA
GGT	1.758 (1.082–2.856)	0.023	3.139 (0.794–12.413)	0.103
CHOL	0.241 (0.067–0.870)	0.030	0.899 (0.832–0.971)	0.018
TG	1.488 (0.320–6.918)	0.613	NA	NA
HDL-C	0.291 (0.079–1.067)	0.063	NA	NA
LDL-C	2.924 (1.699–5.028)	0.001	NA	NA
2D-SWE	6.891 (2.923–16.245)	<0.001	6.847 (1.895–24.733)	0.003

Table 2. Univariate and multivariate logistic regression analysis for Fibrosis-NASH in the training set. CI, confidence interval; NA, not available; OR, odds ratio; SWE, shear wave elastography.

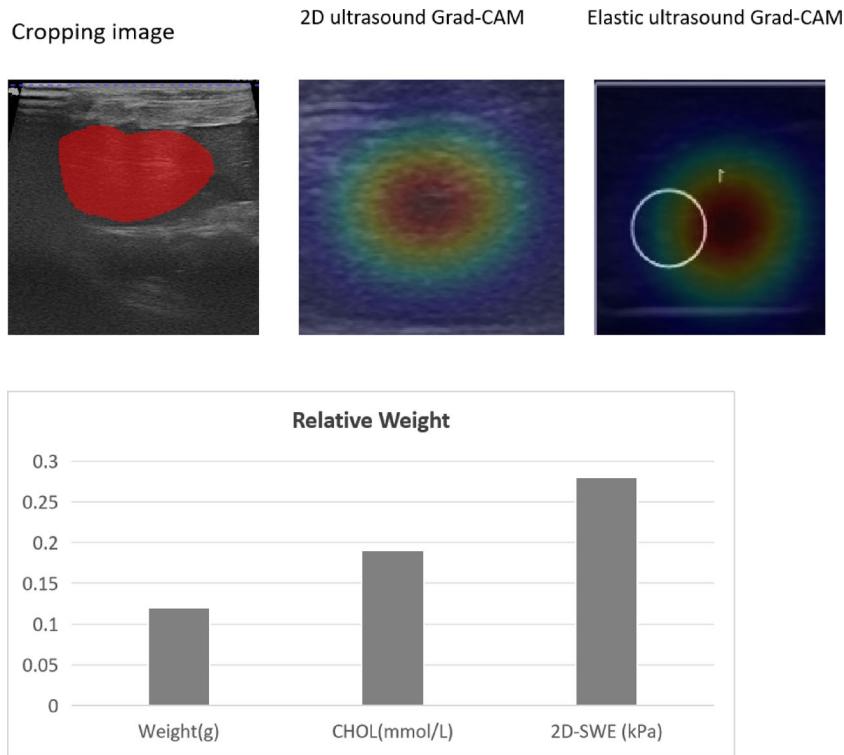


Fig. 3. Representative example of Grad-CAM saliency map from the combined deep learning model. The red-highlighted area indicates the region most relevant to the diagnosis of fibrotic-NASH as identified by the model. The input image was cropped based on the manually delineated Region of Interest (ROI), and Grad-CAM was applied to this cropped image, which corresponds to the actual input used for model prediction.

Fold Description	2D-US Model		SWE Model		Combined Model	
Fold Number	ACC	AUC	ACC	AUC	ACC	AUC
Fold 0	0.793	0.706	0.793	0.699	0.914	0.915
Fold 1	0.776	0.692	0.776	0.768	0.828	0.838
Fold 2	0.759	0.810	0.759	0.860	0.897	0.884
mean \pm SD	0.776 \pm 0.0139	0.736 \pm 0.0526	0.776 \pm 0.0139	0.776 \pm 0.066	0.880 \pm 0.0372	0.879 \pm 0.0316

Table 3. The cross-validation results for different models.

Model Name	Cohort	ACC	AUC	Sen	Spe	PPV	NPV	F1
2D + SWE	TRAIN	0.897 \pm 0.052	0.963 \pm 0.026	0.917 \pm 0.028	0.892 \pm 0.062	0.758 \pm 0.162	0.965 \pm 0.016	0.824 \pm 0.106
2D + SWE	TEST	0.770 \pm 0.027	0.787 \pm 0.103	0.619 \pm 0.206	0.762 \pm 0.226	0.731 \pm 0.345	0.765 \pm 0.118	0.565 \pm 0.143
Clinic	TRAIN	0.782 \pm 0.087	0.788 \pm 0.058	0.677 \pm 0.173	0.826 \pm 0.161	0.881 \pm 0.047	0.619 \pm 0.232	0.618 \pm 0.128
Clinic	TEST	0.782 \pm 0.139	0.727 \pm 0.059	0.585 \pm 0.103	0.827 \pm 0.185	0.803 \pm 0.006	0.708 \pm 0.260	0.613 \pm 0.047
SWE	TRAIN	0.701 \pm 0.198	0.654 \pm 0.193	0.590 \pm 0.085	0.772 \pm 0.231	0.788 \pm 0.082	0.504 \pm 0.353	0.449 \pm 0.299
SWE	TEST	0.746 \pm 0.133	0.515 \pm 0.217	0.463 \pm 0.446	0.856 \pm 0.223	0.834 \pm 0.155	0.389 \pm 0.347	0.404 \pm 0.351

Table 4. Performance metrics of other deep learning models.

Discussion

Multiple studies have confirmed that patients with fibrotic NASH (\geq Stage 2 fibrosis) in NAFLD are at the highest risk of disease progression. Timely diagnosis and treatment of fibrotic NASH can significantly delay disease progression and prevent NAFLD-related complications, which is crucial for improving the long-term quality of life for these patients^{5,21}. Liver biopsy, while the gold standard for monitoring NAFLD progression, has

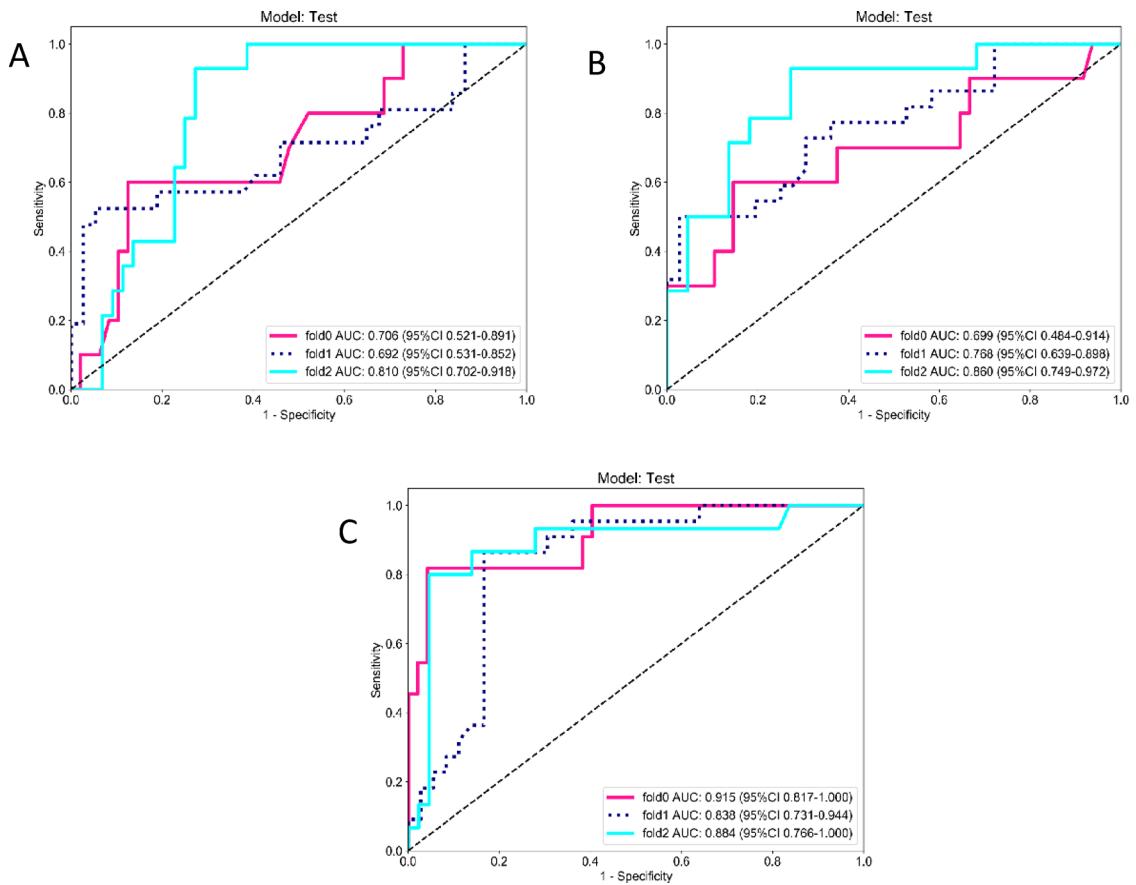


Fig. 4. ROC curves of the three models. (A): ROC curve of the 2 d deep learning model (B): ROC curve of the SWE deep learning model. (C): ROC curve of the combined deep learning model.

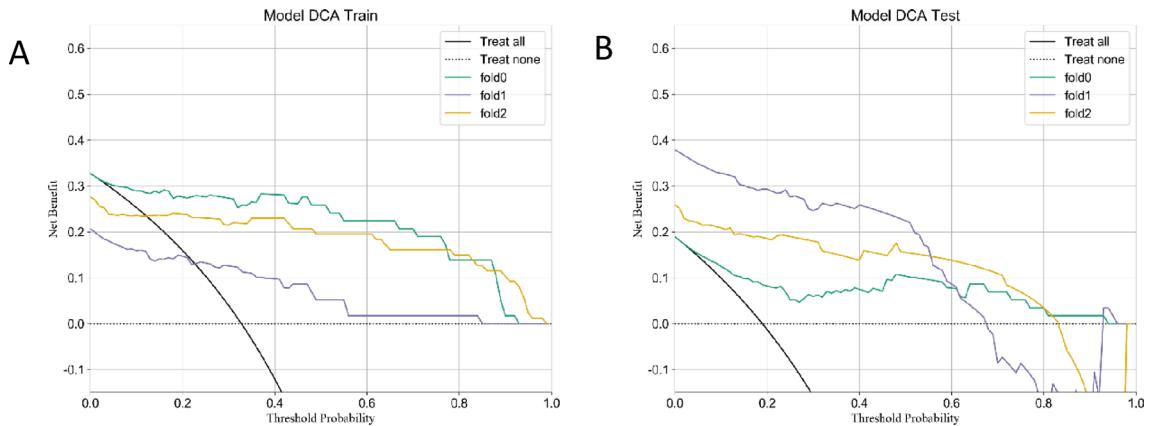


Fig. 5. DCA Curves of the Combined Deep Learning Model A: DCA Curves for Each Cross-Validation in the Training Set B: DCA Curves for Each Cross-Validation in the Validation Set.

certain limitations and drawbacks^{22–24}. Therefore, developing a non-invasive and repeatable method to predict and diagnose the risk level of fibrotic NASH (\geq Stage 2 fibrosis) is crucial.

Currently, deep learning has been widely applied in medical imaging, enabling the extraction of clinically relevant information from complex and diverse clinical datasets, including subtle features often imperceptible to the human eye. This is particularly pertinent to radiological image data, which contains crucial information for disease diagnosis, prognosis, and prediction^{25–28}. In the realm of NAFLD, Omid Pournik et al. developed a predictive neural network model to detect cirrhosis in NAFLD patients. The model achieved a sensitivity of 66%, a false-negative rate of 34%, a specificity of 99%, and a false-positive rate of 1%²⁹. Okanoue T et al. also

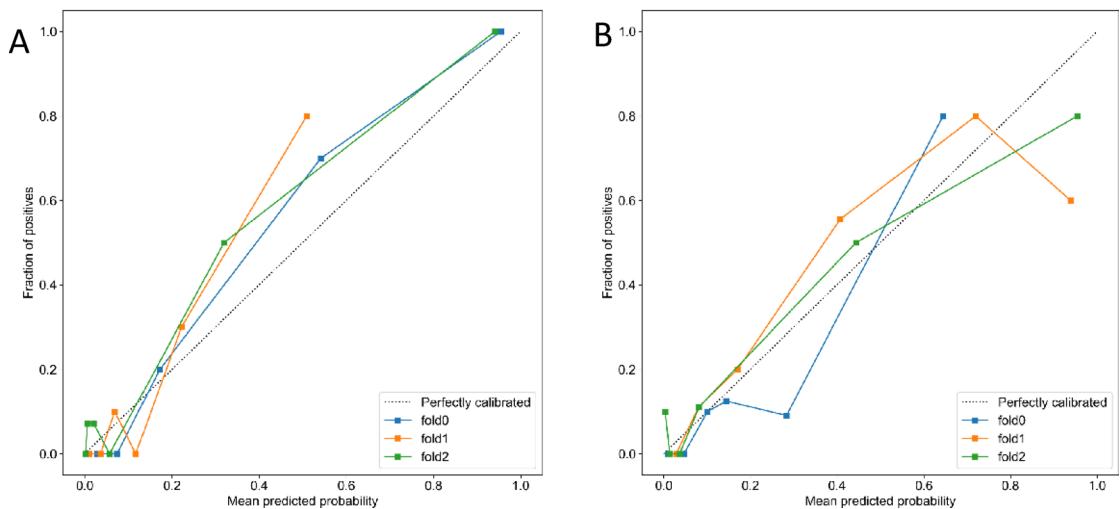


Fig. 6. Calibration curves of the combined deep learning model. **(A)**: Calibration curves for each cross-validation in the training set **(B)**: Calibration curves for each cross-validation in the validation set.

developed a set of high-sensitivity predictive models using artificial neural networks to screen NAFLD and NASH patients, accurately differentiating between mild, moderate, and advanced fibrosis in NASH³⁰. Choi et al. developed a deep learning model using CT images to predict liver fibrosis, achieving AUCs of 0.96, 0.97, and 0.95 for diagnosing F2, F3, and F4 fibrosis stages, respectively³¹. Most relevant to our research topic is the work by Kun W. et al., who developed a deep-learning elastography radiomics predictive model using convolutional neural networks to assess liver fibrosis staging. Their model achieved an AUC of 0.85 (95% CI: 0.81–0.89) for detecting fibrosis stages \geq F2³².

In this study, we employed a NAFLD animal model, enabling the acquisition of comprehensive liver tissue samples for precise pathological diagnosis. This approach circumvents the limitations inherent in patient biopsies, where the small sample size can potentially lead to inaccurate pathological diagnoses³³. We leveraged the convolutional ResNet-18 network, a convolutional neural network architecture incorporating residual structures, which offers several advantages over other deep learning networks. These advantages include the ability to mitigate the degradation problem in deep networks, efficient model complexity suitable for resource-constrained scenarios, and ease of training with reduced susceptibility to the vanishing gradient problem. Even with basic optimizers like SGD, it can achieve good results within fewer training cycles^{26,34,35}. Our study found that the joint deep learning model based on the convolutional ResNet-18 architecture, which integrates 2D ultrasound, SWE, and clinical features, significantly outperformed the deep learning models using 2D ultrasound alone or SWE alone in predicting fibrotic-NASH (\geq F2 fibrosis).

Additionally, we have included in the attachment the results of models based solely on clinical features, SWE values, and 2D + SWE deep learning model, all of which demonstrated inferior performance compared to the combined model. Although several clinical parameters such as body weight, ALT, AST, and SWE showed statistically significant group differences between fibrotic and non-fibrotic NASH in Table 1, models relying on these features alone yielded limited predictive capability. This inconsistency can be explained by several factors. First, statistical significance in univariate group comparisons does not necessarily translate into strong discriminative power for individual-level prediction, as biochemical and metabolic indices often exhibit overlapping distributions across fibrosis stages. Second, SWE measurements capture the mean elasticity within a limited region of interest and are sensitive to factors such as operator pressure, tissue heterogeneity, and inflammatory or steatotic changes, which can obscure fibrosis-specific stiffness variations. Third, clinical indicators primarily reflect systemic metabolic or inflammatory states and fail to capture the spatial distribution and textural heterogeneity of fibrotic lesions that are encoded within imaging modalities. Therefore, single-modality clinical or SWE-based models lack sufficient structural and contextual information to robustly distinguish fibrosis severity. In contrast, the combined 2D + SWE + clinical deep learning model integrates complementary information—linking biochemical context with morphological and biomechanical cues—thus achieving superior accuracy, generalizability, and clinical interpretability.

The combined model achieved a three-fold cross-validation accuracy of 0.880 and an AUC of 0.879, surpassing the performance of a previously reported deep-learning elastography radiomics model. Moreover, the combined model exhibited high specificity, with an average three-fold cross-validation specificity of 0.91. Additionally, the model showed a high positive predictive value (PPV, 0.8) and negative predictive value (NPV, 0.91), indicating its effectiveness in detecting fibrotic NASH cases. This model has the potential to aid clinicians in prioritizing patient evaluation and adopting a more cost-effective approach to assessing NASH status.

This study has several limitations that must be acknowledged. Firstly, the data used in this research were derived from animal models, which may not fully capture the intricate nature and variability of NAFLD in human patients. To enhance the applicability of the findings, it would be advantageous to extend the investigation to encompass a larger and more diverse sample of human participants. The incorporation of human data is crucial

for future research to evaluate the model's performance in clinical settings. Additionally, as this is a single-center study, future research should utilize independent datasets from various sources or populations to conduct external validation and assess the model's robustness and generalizability.

Conclusions

In this study, we developed and validated a novel multimodal deep learning model, ResHy-Net, integrating 2D ultrasound, real-time shear wave elastography (SWE) images, and key clinical features to predict fibrotic nonalcoholic steatohepatitis (fibrotic-NASH, $\geq F2$). In three-fold cross-validation, the combined model achieved a mean accuracy (ACC) of 0.880 and a mean area under the receiver operating characteristic curve (AUC) of 0.879, significantly outperforming the single-modality models using only 2D ultrasound (ACC 0.776, AUC 0.736) or only SWE (ACC 0.776, AUC 0.776), demonstrating that multimodal information fusion substantially enhances predictive accuracy and discriminative power. Further evaluation across sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) showed that the combined model attained an average specificity of 0.91, PPV of 0.80, and NPV of 0.91, indicating robust performance in minimizing false negatives and false positives. Grad-CAM visualization revealed that the model automatically focuses on liver tissue regions associated with significant fibrosis and that clinical variables (e.g., body weight, total cholesterol, and SWE values) contribute substantially to its decisions, confirming the interpretability and clinical relevance of multimodal feature fusion. Clinically, a noninvasive, repeatable tool for fibrotic-NASH screening is crucial for early diagnosis and timely intervention. The ResHy-Net model not only demonstrated efficacy in animal experiments but also lays a technological foundation for subsequent clinical deployment. By seamlessly integrating with routine B-mode and SWE ultrasound workflows, the model can provide real-time risk assessment on ultrasound workstations, assisting physicians in accurately identifying high-risk patients requiring pharmacological intervention, thereby optimizing care pathways and reducing reliance on liver biopsy. This study's limitations include its reliance on an SD rat model, necessitating external validation in multicenter human clinical cohorts. Additionally, the current model does not incorporate serum biochemical panels or genomic data; future work could explore additional modalities to enhance generalizability and individualized prediction accuracy. Next steps will focus on building a large-scale, multicenter clinical sample database, conducting prospective studies and cost-effectiveness analyses to advance the model toward clinical translation.

In summary, the proposed multimodal deep learning model exhibits outstanding performance in predicting fibrotic-NASH and offers a powerful tool for early identification and clinical decision-making in nonalcoholic steatohepatitis patients, with promising prospects for broad clinical application.

Data availability

Data is provided within the manuscript or supplementary information files. Data types and formats: - The types of data include raw experimental data, processed data. - Data are available in Excel formats. Supporting materials: - Additional supporting materials such as codebooks and analysis scripts are available from the corresponding author upon request. Contact information: - For any queries or further information, please contact Dr. Wang at wjl980134@163.com.

Received: 13 December 2024; Accepted: 12 November 2025

Published online: 29 December 2025

References

- Ravaoli, F. et al. Diagnostic accuracy of FibroScan-AST (FAST) score for the non-invasive identification of patients with fibrotic non-alcoholic steatohepatitis: a systematic review and meta-analysis. *Gut* **72** (7), 1399–1409. <https://doi.org/10.1136/gutjnl-2022-328689> (2023).
- Velliou, R. I., Legaki, A. I., Nikolakopoulou, P., Vlachogiannis, N. I. & Chatzigeorgiou, A. Liver endothelial cells in NAFLD and transition to NASH and HCC. *Cell Mol Life Sci.* **80**(11):314. Published 2023 Oct 5. (2023). <https://doi.org/10.1007/s0018-023-04966-7>
- Aggarwal, M. et al. Development of machine learning model to detect fibrotic non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. *Dig. Liver Dis.* **53** (12), 1669–1672. <https://doi.org/10.1016/j.dld.2021.07.016> (2021).
- Rowe, I. A., Wong, V. W. & Loomba, R. Treatment candidacy for Pharmacologic therapies for NASH. *Clin. Gastroenterol. Hepatol.* **20** (6), 1209–1217. <https://doi.org/10.1016/j.cgh.2021.03.005> (2022).
- Taylor, R. S. et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology* **158** (6), 1611–1625e12. <https://doi.org/10.1053/j.gastro.2020.01.043> (2020).
- Pinzani, P. M. Liver fibrosis in NAFLD/NASH: from pathophysiology towards diagnostic and therapeutic strategies. *Mol. Aspects Med.* **95**, 101231. <https://doi.org/10.1016/j.mam.2023.101231> (2024).
- Powell, E. E., Wong, V. W. & Rinella, M. Non-alcoholic fatty liver disease. *Lancet* **397** (10290), 2212–2224. [https://doi.org/10.1016/S0140-6736\(20\)32511-3](https://doi.org/10.1016/S0140-6736(20)32511-3) (2021).
- Chuah, K. H. et al. MACK-3 (combination of hoMa, Ast and CK18): A promising novel biomarker for fibrotic non-alcoholic steatohepatitis. *Liver Int.* **39** (7), 1315–1324. <https://doi.org/10.1111/liv.14084> (2019).
- Newsome, P. N. et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study [published correction appears in Lancet Gastroenterol Hepatol. 5(4):e3. doi: 10.1016/S2468-1253(20)30055-8]. *Lancet Gastroenterol Hepatol.* **2020**;5(4):362–373. (2020). [https://doi.org/10.1016/S2468-1253\(19\)30383-8](https://doi.org/10.1016/S2468-1253(19)30383-8)
- Kang, B. K., Kim, M., Song, S. Y., Jun, D. W. & Jang, K. Feasibility of modified Dixon MRI techniques for hepatic fat quantification in hepatic disorders: validation with MRS and histology. *Br. J. Radiol.* **91** (1089), 20170378 (2018).
- Cao, L. L. et al. Dietrich C.F. Artificial intelligence in liver ultrasound. *World J. Gastroenterol.* **28** (27), 3398–3409. <https://doi.org/10.3748/wjg.v28.i27.3398> (2022).
- Raufy, M. R. et al. A novel method for diagnosing cirrhosis in patients with chronic hepatitis B: artificial neural network approach. *J. Med. Syst.* **35** (1), 121–126. <https://doi.org/10.1007/s10916-009-9348-8> (2011).
- Kriegeskorte, N. & Golan, T. Neural network models and deep learning. *Curr. Biol.* **29** (7), R231–R236. <https://doi.org/10.1016/j.cub.2019.02.034> (2019).

14. Sato, M. et al. Development of novel deep multimodal representation learning-based model for the differentiation of liver tumors on B-mode ultrasound images. *J. Gastroenterol. Hepatol.* **37** (4), 678–684. <https://doi.org/10.1111/jgh.15763> (2022).
15. Zhang, C. et al. A deep learning-powered diagnostic model for acute pancreatitis. *BMC Med. Imaging.* **24** (1), 154. <https://doi.org/10.1186/s12880-024-01339-9> (2024). Published 2024 Jun 20.
16. Hong, J. et al. A novel hierarchical deep learning framework for diagnosing multiple visual impairment diseases in the clinical environment. *Front. Med. (Lausanne)*. **8**, 654696. <https://doi.org/10.3389/fmed.2021.654696> (2021). Published 2021 Jun 7.
17. Kudo, M. et al. JSUM ultrasound elastography practice guidelines: liver. *J Med Ultrason* **2013**;40(4):325–357. (2001). <https://doi.org/10.1007/s10396-013-0460-516>
18. Giovanna Ferraioli, Carmine Tinelli, Barbara Dal Bello, Mabel Zicchetti, Gaetano Filice, Carlo Filice. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology* **56** (6), 2125–2133. <https://doi.org/10.1002/hep.25936> (2012).
19. Anthony, E. S. et al. Shear-wave elastography for the Estimation of liver fibrosis in chronic liver disease: determining accuracy and ideal site for measurement. *Radiology* **274** (3), 888–896. <https://doi.org/10.1148/radiol.14140839> (2015).
20. Groen, A. M., Kraan, R., Amirkhan, S. F., Daams, J. G. & Maas, M. A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: limited use of explainable AI? *Eur. J. Radiol.* **157**, 110592. <https://doi.org/10.1016/j.ejrad.2022.110592> (2022).
21. Ajmera, V. & Loomba, R. Imaging biomarkers of NAFLD, NASH, and fibrosis. *Mol. Metab.* **50**, 101167. <https://doi.org/10.1016/j.molmet.2021.101167> (2021).
22. Tarantino, G. Is assessing the presence of NASH by liver histology or surrogate markers always advisable? *Hepat. Mon.* **13** (2), e7560. <https://doi.org/10.5812/hepatmon.7560> (2013). Published 2013 Feb 20.
23. Houghton, D. et al. The degree of hepatic steatosis associates with impaired cardiac and autonomic function. *J. Hepatol.* **70** (6), 1203–1213. <https://doi.org/10.1016/j.jhep.2019.01.035> (2019).
24. Brunt, E. M. Pathology of fatty liver disease. *Mod. Pathol.* **20** (Suppl 1), S40–S48. <https://doi.org/10.1038/modpathol.3800680> (2007).
25. Havaei, M. et al. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31. <https://doi.org/10.1016/j.media.2016.05.004> (2017).
26. Zhou, Q. et al. Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning. *Curr. Pharm. Des.* **28** (4), 287–295. <https://doi.org/10.2174/138161282766211213143357> (2022).
27. Nam, D., Chapiro, J., Paradis, V., Seraphin, T. P. & Kather, J. N. Artificial intelligence in liver diseases: improving diagnostics, prognostics and response prediction. *JHEP Rep.* **4** (4), 100443. <https://doi.org/10.1016/j.jhepr.2022.100443> (2022). Published 2022 Feb 2.
28. Rezazade Mehrizi, M. H., van Ooijen, P. & Homan, M. Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. *Eur. Radiol.* **31** (4), 1805–1811. <https://doi.org/10.1007/s00330-020-07230-9> (2021).
29. Pournik, O., Dorri, S., Zabolinezhad, H., AlavianSM & Eslami, S. A diagnostic model for cirrhosis in patients with non-alcoholic fatty liver disease: an artificial neural network approach. *Med. J. Islam Repub. Iran.* **28**, 116 (2014).
30. Takeshi, O., Toshihide Shima, Yasuhide Mitsumoto, Atsushi Umemura, Kanji Yamaguchi, Yoshito Itoh, Masato Yoneda, Atsushi Nakajima, Eishiro Mizukoshi, Shuichi Kaneko, Kenichi Harada. Novel artificial intelligent/neural network system for staging of nonalcoholic steatohepatitis. *Hepatol. Res.* **51** (10), 1044–1057. <https://doi.org/10.1111/hepr.13681> (2021).
31. Kyu Jin Choi, Jong Keon Jang, Seung Soo Lee, Yu sub Sung, Woo Hyun Shim, Ho Sung Kim, Jessica Yun, Jin-Young Choi, Yedaun Lee, Bo-Kyeong Kang, Jin Hee Kim, So Yeon Kim, Eun Sil Yu. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver. *Radiology* **289**(3):688–697. (2018). <https://doi.org/10.1148/radiol.2018180763>
32. Wang, K. 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* **68** (4), 729–741. <https://doi.org/10.1136/gutjnl-2018-316204> (2019).
33. Castera, L. & Pawlotsky, J. M. Noninvasive diagnosis of liver fibrosis in patients with chronic hepatitis C. *MedGenMed* **7** (4), 39 (2005). Published 2005 Nov 9.
34. Liu, Y., She, G. R. & Chen, S. X. Magnetic resonance image diagnosis of femoral head necrosis based on ResNet18 network. *Comput. Methods Programs Biomed.* **208**, 106254. <https://doi.org/10.1016/j.cmpb.2021.106254> (2021).
35. Yuliana & Jiménez-Gaona , María José Rodríguez Álvarez, Darwin Castillo-Malla, Santiago García-Jaen, Diana Carrón-Figueroa, Patricio Corral-Domínguez, Vasudevan Lakshminarayanan. BraNet: a mobil application for breast image classification based on deep learning algorithms. *Med. Biol. Eng. Comput.* **62** (9), 2737–2756. <https://doi.org/10.1007/s11517-024-03084-1> (2024).

Author contributions

Author Contributions: (I) Conception and design: F. X.; (II) Provision of study materials : J.W. and C.Z. (III) Collection and assembly of data: K.W. and F.X.; (IV) Data analysis and interpretation: F.X. and Y.W.; (V) Manuscript writing: F.X.; (VI) Final approval of manuscript: All authors. F.X. and K.W. have contributed equally to this work.

Funding

This work was supported by the institutional research project fund of the Second People's Hospital of Wuhu City, Anhui Province (JC2023B05).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

All animal experiments conducted in this study were approved by the Institutional Animal Care and Use Committee of the Second People's Hospital of Wuhu, Anhui Province (Approval No. 2024-KY-112).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-28753-5>.

Correspondence and requests for materials should be addressed to J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025