Journal of Liver Transplantation 21 (2026) 100318

Contents lists available at ScienceDirect

Journal of Liver Transplantation

journal homepage: www.elsevier.com/locate/liver

Review

Deep learning for non-invasive NAFLD detection and staging: A comprehensive review

Priyanka Sengar *, Jagendra Singh, Abhay Bansal

School of Computer Science Engineering and Technology, Bennett University, Greater Noida 201310, Uttar Pradesh, India

ARTICLE INFO

Keywords:
Non-alcoholic fatty liver disease
Deep learning
Ultrasound imaging
Disease staging
Explainable AI
Non-invasive diagnostics

ABSTRACT

Non-alcoholic fatty liver disease (NAFLD), a prevalent chronic liver condition, demands accurate, non-invasive diagnostics to replace invasive liver biopsies for staging steatosis, non-alcoholic steatohepatitis (NASH), and fibrosis. Deep learning (DL) has demonstrated transformative potential in enhancing diagnostic accuracy and efficiency by leveraging ultrasound (US) imaging, elastography, and clinical/serological data. This systematic review analyzes 64 studies from 2015 to 2025, retrieved from multiple scholarly databases, to evaluate DL models for NAFLD detection and quantification. The reviewed models, primarily leveraging convolutional neural networks (CNNs) and multimodal data integration, achieve high diagnostic accuracy (AUC $>$ 0.90) and generalizability in detecting and staging NAFLD. Ablation studies highlight the critical role of multimodal inputs and advanced architectures in improving performance. However, gaps such as limited diverse datasets, scarce prospective validations, and poor model explainability persist. Opportunities include developing explainable AI (XAI), federated learning for multi-institutional collaboration, and integration with telemedicine for scalable diagnostics. These findings suggest that DL-based systems can significantly reduce biopsy dependency, enhance early detection, and improve clinical outcomes, provided interdisciplinary efforts address existing challenges.

1. Introduction

Chronic liver disease (CLD), particularly NAFLD, has become a growing global health concern, affecting nearly one-third of the adult population worldwide [1]. NAFLD is characterized by the accumulation of fat in hepatocytes in individuals without a history of significant alcohol consumption. It represents a hepatic manifestation of metabolic syndrome and is often associated with conditions like insulin resistance, obesity, and type 2 diabetes [2,3]. The disease spectrum ranges from simple steatosis to non-alcoholic steatohepatitis (NASH), which involves hepatic inflammation and cellular injury, and can progress to fibrosis, cirrhosis, liver failure, or hepatocellular carcinoma. While lifestyle interventions can reverse early-stage NAFLD, advanced forms of the disease may lead to irreversible organ damage, necessitating liver transplantation [4,5]. Given the progressive nature of NAFLD, early detection and accurate staging are critical for effective clinical management. Differentiating between steatosis, NASH, and fibrosis stages helps inform prognostic decisions and treatment planning. Various non-invasive approaches have been developed to aid in diagnosis, including clinical scoring indices (e.g., SteatoTest, NAFLD Fibrosis Score, Fibrosis-4 Index) and imaging modalities like ultrasound (US), US elastography, computed tomography (CT), and magnetic resonance imaging (MRI) [6]. Despite their accessibility, conventional US methods remain subjective and operator dependent [7]. Elastography and MRI provide more quantitative assessments, but their accuracy can be limited by technical constraints and cost [8]. Liver biopsy, while definitive, is invasive and not feasible for routine screening, restricting its use to select cases with high clinical suspicion [9]. In recent years, artificial intelligence (AI), particularly DL techniques, has emerged as a transformative tool in medical imaging and diagnostics. AI algorithms can process large, heterogeneous datasets and uncover complex patterns that may not be discernible through conventional analytical methods [10]. In the context of NAFLD, AI systems have been explored for integrating clinical, serological, and imaging data to improve diagnostic accuracy and reduce dependency on invasive procedures. While several reviews have discussed AI applications in liver disease diagnosis [10], they often lack focus on the specific contributions of input data types (e.g., ultrasound vs. clinical) and the progression stages of NAFLD/NASH. Moreover, few reviews offer a side-by-side comparative assessment of AI model performance across different modalities. This review addresses

* Corresponding author.
E-mail address: priyanka06mps@gmail.com (P. Sengar).

https://doi.org/10.1016/j.liver.2026.100318
Received 4 September 2025; Received in revised form 31 December 2025; Accepted 6 January 2026
Available online 16 January 2026

these gaps by systematically examining DL-based diagnostic models developed between 2015 and 2025 for detecting and quantifying NAFLD and its complications. We focus specifically on models that utilize ultrasound (including elastography) and clinical datasets, given their cost-effectiveness, non-invasiveness, and clinical applicability in high-risk populations. Our objective is to critically evaluate the performance of these models, assess their potential for integration as clinical decision-support tools, and provide practical guidance for researchers and clinicians involved in developing automated diagnostic frameworks. By examining the relationships between various data types and disease stages, this review offers a comprehensive perspective on the role of AI in enhancing the detection and quantification of NAFLD. In doing so, we not only analyze the current state of AI applications in NAFLD diagnostics but also highlight key trends, challenges, and future directions. This paper aims to serve as a foundational resource for healthcare practitioners, AI developers, and biomedical researchers by offering actionable insights into model design, input data selection, and disease staging. The findings highlight the potential of AI-enabled systems to augment clinical workflows, facilitate early detection, and ultimately enhance patient outcomes for those with chronic liver disease [11].

The main contribution of the paper is given as follows:

- This paper systematically reviews DL-based diagnostic models developed over the last decade, focusing on their application to non-alcoholic fatty liver disease (NAFLD) detection and staging.
- The review distinctly categorizes and compares models based on two primary input types, ultrasound imaging (including elastography) and clinical/serological parameters, highlighting their respective advantages, limitations, and performance trends.
- The paper evaluates how AI models perform across different pathological stages of NAFLD, including steatosis, steatohepatitis (NASH), and fibrosis, providing clarity on model suitability for specific clinical needs.
- It assesses the diagnostic accuracy, efficiency, and clinical applicability of reviewed models, offering guidance on their potential integration as non-invasive tools in routine medical practice.
- The study identifies key limitations in existing literature and proposes future research directions to enhance the development of robust, generalizable AI-based diagnostic systems for chronic liver disease.

The organization of the paper is structured as follows: Section 2 details the research methodology, adopting Kitchenham's guidelines to formulate research questions, develop a search strategy across eight electronic databases, select 64 studies from 2015 to 2025, and synthesize data to address gaps in DL for NAFLD diagnosis. Section 3 presents the results and discussion, addressing five research questions. Section 4 reviews open-source databases for NAFLD research, highlighting their role in facilitating data-driven studies. Section 5 explores future directions, proposing lightweight DL architectures, multimodal data integration, XAI, and federated learning to enhance clinical applicability. Section 6, the ablation study, synthesizes findings from studies comparing input modalities model architectures and preprocessing techniques, demonstrating their impact on diagnostic performance. Section 7, the conclusion, summarizes the high accuracy of DL models (>90 % AUC), identifies gaps like limited prospective validations, and underscores opportunities for non-invasive, scalable NAFLD diagnostics through interdisciplinary collaboration and regulatory frameworks.

2. Research method

Kitchenham's guidelines [12] have been adopted as the research methodology to explore key issues in the domain of AI-driven learning models for diabetes detection. The following essential steps are involved in this process: formulating research questions (RQs), developing a robust search strategy, selecting relevant studies, and synthesizing the collected data.

2.1. Research questions

The goal of this SLR is to uncover research gaps in the field of DL Models for Chronic Liver Disease Detection and Quantification. To address this objective, the following research questions have been carefully developed as part of the SLR process:

RQ1. How effective are AI-based ultrasound techniques in the quantitative assessment and classification of NAFLD, including liver steatosis, fibrosis, and inflammation?
RQ2. What is the optimal combination of clinical and serological features for machine learning models to accurately diagnose different stages of NAFLD (e.g., simple steatosis, NASH, fibrosis) compared to using clinical features alone?
RQ3. To what extent have DL models been successful in distinguishing between different pathological stages of NAFLD, such as steatosis, NASH, and fibrosis?
RQ4. What are the common challenges and limitations encountered in developing and deploying DL models for chronic liver disease diagnosis in clinical settings?
RQ5. What trends, gaps, and opportunities exist in the literature for enhancing the integration of DL approaches in non-invasive chronic liver disease diagnostics?

2.2. Search strategy

A search string has been formulated using key terms from the research questions to extract primary studies from the following eight electronic databases: ACM Digital Library, arXiv.org, IEEE Xplore, PLOS, ScienceDirect, SpringerLink, and Wiley Online Library.

Search string: *((AI OR "artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "convolutional neural network" OR "transformer model") AND (method OR approach OR technique OR framework OR model OR architecture OR algorithm) AND (detection OR prediction OR diagnosis OR identification OR classification OR quantification OR staging) AND ("chronic liver disease" OR NAFLD OR "non-alcoholic fatty liver disease" OR NASH OR "non-alcoholic steatohepatitis" OR fibrosis OR steatosis))*

2.3. Study selection process

In this work, data have been extracted from studies published over the last eleven years (i.e., from 2015 to 2025). The complete selection process for the studies is illustrated in Fig. 1. In the initial search process, 1370 studies were identified from the eight different digital database resources: 326 studies on the ACM Digital Library, 78 studies on arXiv.org, 204 studies on IEEE Xplore, 11 studies on PLOS, 525 studies on ScienceDirect, 131 studies on SpringerLink, and 95 studies on the Wiley Online Library.

To address the five research questions comprehensively, we conducted a systematic literature review of published studies from 2015 to 2025. The initial screening involved a thorough search of eight major scholarly databases: ACM Digital Library, arXiv.org, IEEE Xplore, PLOS, ScienceDirect, SpringerLink, and Wiley Online Library using a combination of relevant keywords and Boolean operators. A total of 1370 studies were initially retrieved. After removing duplicates and clearly irrelevant entries based on title review, 487 studies were shortlisted as primary studies. From these 487 primary studies, further filtering was performed based on an analysis of the abstract and conclusion, resulting in the selection of 291 studies. To ensure completeness, the references of these 185 studies were also manually reviewed, leading to the identification of 98 additional relevant studies. This brought the total number of selected studies to 64, which were then subjected to a rigorous quality

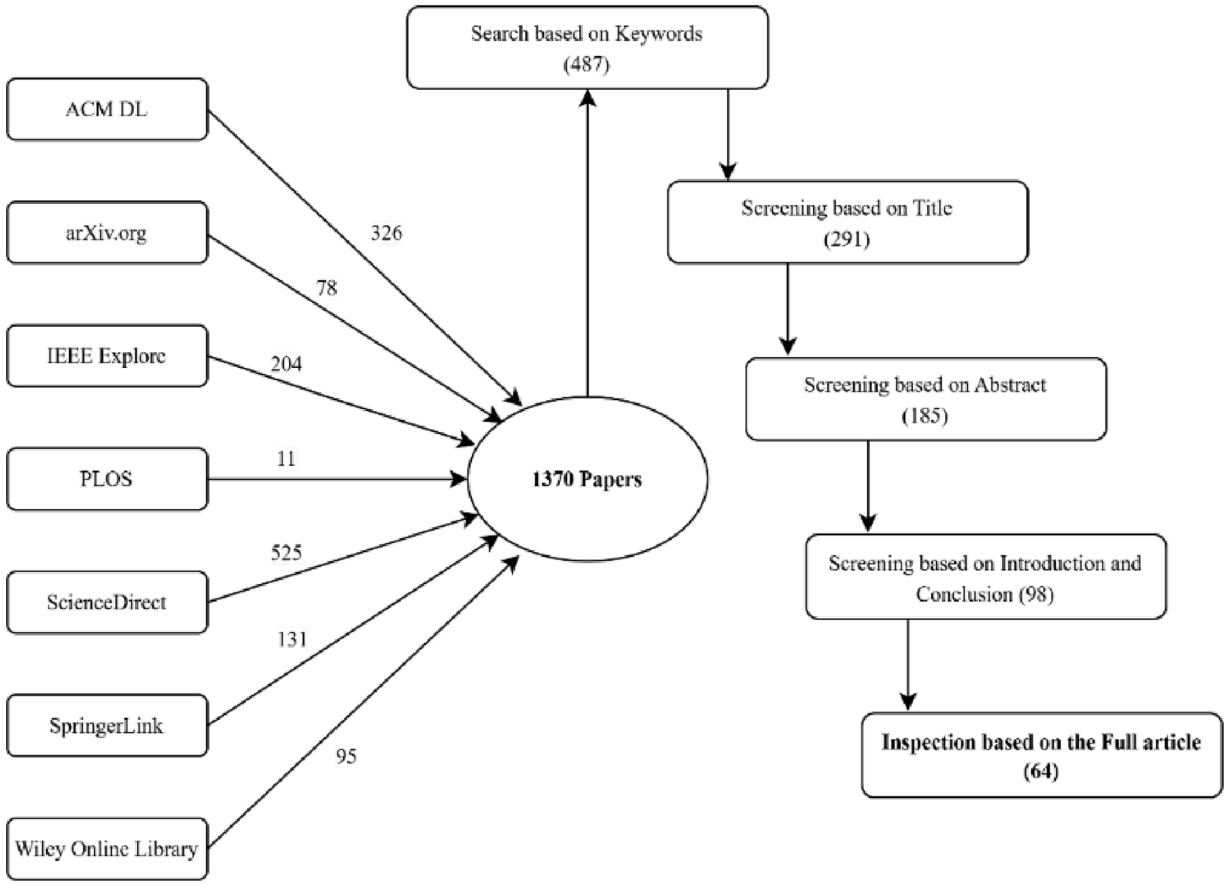

Fig. 1. Study's selection process (PRISMA).

assessment (QA) process.

The QA process aimed to ensure the relevance, credibility, and methodological soundness of the selected studies. The following assessment questions were applied:

- QA1: Is the sample size adequate to draw reliable conclusions?
- QA2: Are the methods used well described and appropriate for DL applications in CLD diagnosis?
- QA3: Are the results clearly presented and statistically validated?
- QA4: Are the limitations of the proposed methods and models critically discussed?
- QA5: Are there innovative DL architectures or frameworks proposed for the detection or quantification of NAFLD stages, and is their clinical integration addressed?

Each QA question was scored as follows: (a) Yes = 1, (b) Partly = 0.5, and (c) No = 0.

Only studies scoring ≥3.5 out of 5 (See Annexure A and Annexure B in See Supplementary Material File) were included for final analysis to address the following research questions:

- RQ1: To explore the AI-based ultrasound techniques in the quantitative assessment and classification of NAFLD.
- RQ2: To compare the performance and clinical viability of models using ultrasound imaging versus those using clinical/serological data.
- RQ3: To evaluate the effectiveness of DL models in distinguishing between pathological stages such as steatosis, NASH, and fibrosis.
- RQ4: To identify challenges and limitations in model development and deployment in clinical workflows.
- RQ5: To map out the emerging trends, gaps, and opportunities for improving the integration of DL techniques in non-invasive diagnostics of chronic liver disease.

2.4. Inclusion and exclusion criteria

Table 1 (See Supplementary Material) summarizes the criteria used for including and excluding studies in the systematic review, as described in the provided document. It highlights the focus on non-invasive AI/DL applications for NAFLD diagnosis using ultrasound and/or clinical data within a specific time frame and quality threshold.

2.5. Threats to validity

We created a search string to identify primary studies from eight different digital libraries. Practically, it is not possible to extract all studies from the terms that are used in the RQs. To counter this threat, a manual scrutiny of the studies was conducted to identify the relevant primary studies that might have been overlooked in the initial search phase of the SLR. To minimise the inaccuracy of the retrieved data, the quality assessment of the studies was conducted independently.

3. Results and discussion

This section's goal is to discuss the formulated RQs based on the 64 selected studies. Fig. 2 (See Supplementary Material) shows the year-wise distribution of the 64 selected studies.

Table 1
Inclusion and exclusion criteria for study selection.

Category	Criteria	Rationale
Inclusion Criteria	Focus on Non-Invasive Diagnosis: Studies that utilized AI/DL models for the non-invasive detection, classification, or quantification of chronic liver diseases, specifically NAFLD and its stages (steatosis, NASH, fibrosis).	To ensure the review focuses on the application of AI/DL for the intended diagnostic purpose, aligning with the research questions.
	Use of Ultrasound and/or Clinical Data: Studies that employed ultrasound imaging (including its variants like elastography) and/or clinical/serological data as input for the AI/DL models.	To align with the scope mentioned in the abstract and introduction, the focus is on commonly used non-invasive modalities for assessing liver disease.
	Publication Period: Studies published between January 2015 and December 2025.	To capture the recent advancements in the field within a defined and relevant timeframe.
	Peer-Reviewed Publications: Original research articles published in peer-reviewed journals, conference proceedings, or pre-print servers (e.g., arXiv) that met a quality assessment score of ≥ 3.5.	To ensure the inclusion of scientifically rigorous and relevant studies that have undergone some form of scrutiny. The quality assessment threshold aimed to filter out low-quality studies.
	English Language: Studies published in the English language.	Due to resource constraints and the common language of scientific communication in the field.
	Clear Methodology and Results: Studies that provided sufficient detail on the AI/DL model architecture, input data, experimental setup, and reported quantitative performance metrics (e.g., accuracy, AUC, sensitivity, specificity).	To allow for meaningful analysis, comparison, and synthesis of the findings across different studies.
Exclusion Criteria	Invasive Diagnostic Methods: Studies primarily focused on AI/DL analysis of data from invasive procedures (e.g., liver biopsy histopathology) without integration with non-invasive modalities within the scope.	To maintain the focus on non-invasive diagnostic approaches.
	Animal or In-vitro Studies: Studies conducted on animal models or in-vitro experiments without direct application to human chronic liver disease diagnosis using non-invasive methods.	To ensure the relevance of the reviewed literature to clinical diagnostics in humans.
	Reviews, Editorials, and Letters: Review articles, editorials, letters to the editor, and similar publication types that did not present original research findings.	To focus on primary research that contributed novel AI/DL models or applications in the field.
	Studies with Insufficient Information: Studies lacking sufficient detail on the AI/DL methodology, data characteristics, or performance metrics that would prevent meaningful analysis and comparison.	To ensure that only studies with enough information for quality synthesis were included.
	Non-English Language Publications: Studies published in languages other than English.	Due to resource constraints for translation and analysis.
	Low Quality Assessment Score: Studies that did not meet the minimum quality assessment score of 3.5.	To exclude studies deemed to be of lower methodological quality, which potentially affects the reliability of their findings.
	Focus on Other Liver Diseases: Studies primarily focused on other liver diseases (e.g., viral hepatitis, liver cancer) without a significant component addressing NAFLD detection or staging using non-invasive AI/DL methods within the specified data modalities.	To ensure the review remains focused on NAFLD and its related conditions, as defined in the research scope.

RQ1. How effective are AI-based ultrasound techniques in the quantitative assessment and classification of NAFLD, including liver steatosis, fibrosis, and inflammation?

Ultrasound (US) imaging is a widely accessible, cost-effective, and routinely used modality for evaluating liver steatosis. In standard US, linear sound wave pulses propagate through tissue, and due to nonlinear interactions, reflected signals contain harmonic frequencies. Clinical systems typically use second harmonic echoes to enhance image quality by improving signal-to-noise ratio, reducing artifacts, and sharpening tissue boundaries. While liver fat increases echogenicity in US images, conventional assessments remain qualitative and have limited sensitivity, particularly for detecting mild steatosis. Advanced US techniques, such as elastography, enhance diagnostic capabilities by applying acoustic radiation force impulses to assess liver stiffness, which correlates with the severity of fibrosis. This is achieved by tracking the velocity of shear waves generated in the liver and converting the measurements into kilopascals—an approach implemented in tools like FibroScan [13]. Recent literature highlights the integration of artificial intelligence (AI) with US imaging, primarily through CNNs, to automate and improve the diagnosis of liver conditions. These models have been utilised for binary classification (normal vs. abnormal) [14], staging of steatosis [15], fibrosis [16], NAFLD grading [17], and inflammation [15]. For instance, Andrade et al. developed a classification system combining neural networks, support vector machines (SVM), and k-nearest neighbors (KNN), achieving an accuracy of 79.8 % [18]. Destrempes et al. applied random forest (RF) algorithms with bootstrapping to analyze quantitative US and elastography images from 82 liver samples, reporting C-statistics of 90 %, 77 %, and 75 % for steatosis, fibrosis, and inflammation grading, respectively, surpassing elastography alone [15]. Gaber et al. enhanced classification using radiomics-based feature extraction combined with a voting classifier built from multiple extreme learning machines and majority voting [19]. Meanwhile, Byra et al. employed a pre-trained Inception-ResNet DL model on US images from bariatric surgery patients, using extracted features with an SVM for classification. Their method outperformed traditional texture analysis (e.g., gray-level co-occurrence matrix) and manual grading using the hepatorenal index [20].

Table 2 summarizes various AI algorithms applied to ultrasound and elastography data for non-invasive liver disease assessment, including steatosis, fibrosis, and NAFLD classification and staging. Different models, ranging from traditional machine learning like Random Forests and Voting Classifiers to DL architectures such as CNNs, ResNet variants, VGG-19, and Inception-V3, have been employed on datasets with varying sample sizes. The reported performance metrics, primarily accuracy and AUC, demonstrate promising results across various classification targets and modalities. Some studies highlight the impact of ROI selection, ensemble learning, image transformations, and the influence of ultrasound acquisition settings on model performance.

RQ2. What is the optimal combination of clinical and serological features for machine learning models to accurately diagnose different stages of NAFLD compared to using clinical features alone?

The accurate diagnosis and staging of NAFLD remain significant challenges in clinical hepatology, especially considering the silent

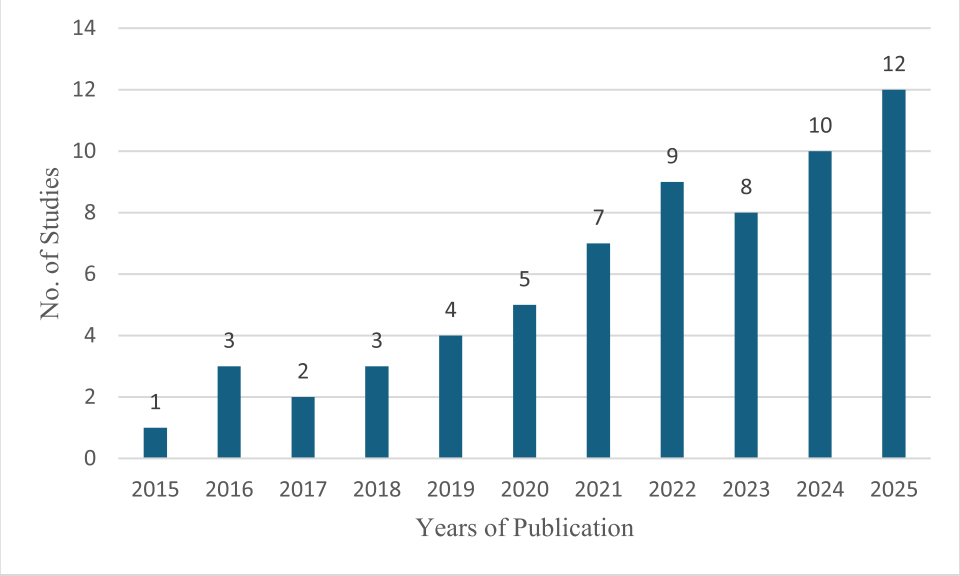

Fig. 2. Year-wise distribution of the selected studies.

Table 2
AI algorithm performance in non-invasive liver disease assessment (Ultrasound).

Ref.	AI Algorithm	Modality	Sample Size	Classification Target	Performance	Key Notes
[15]	RF + Bootstrapping	Quantitative US + Elastography	82 patients	Steatosis, Fibrosis, Inflammation	AUC: 0.9 (S), 0.77 (F), 0.75 (I)	Multi-label grading using RF
[19]	Voting Classifier	US	300 images	Normal vs Fatty Liver	Acc: 0.957, F1: 0.956	Genetic algorithm for ROI selection, radiomics
[21]	DL Ensemble + SVM	B-mode	55 participants	Steatosis stage	Acc: 0.986, Spec: 1.0	Ensemble learning using pretrained CNNs
[20]	ResNet-50	US	135 participants	Fatty liver classification	AUC: 0.91	Regression + classification from multi-view US
[17]	CNN	US	240 participants	NAFLD severity (4 classes)	AUC: 0.958	Grayscale + envelope signal analysis
[22]	Multi-scale CNN	US	55 patients	Healthy vs Fatty liver	Acc: 0.978, AUC: 1.0	Fused image transformations improve results
[23]	ResNet-50 V2	B-mode	2070 patients (21,855 images)	Steatosis stages	Acc: 0.841, Spec: 0.948	Large cohort and multiclass classification
[24]	CNN (Inception-V3)	B-mode	60 patients	Normal vs Steatosis	Acc: 0.932, AUC: 0.93	DL on grayscale B-mode US
[25]	CNN	US	204 patients	NAFLD vs No NAFLD	Acc: 0.96, Spec: 0.94	Model impacted by ultrasound acquisition settings
[26]	VGG-19 (pretrained)	B-mode	90 subjects	Fatty liver detection	Acc: 0.801, Prec: 0.862	Multi-view US approach (liver + kidney)
[27]	ResNet-18	US	3310 patients	Steatosis stages	Acc: 0.87	Outperformed FibroScan
[28]	Cascaded CNN	US	512 images	Steatosis stages	Acc: 0.999, Spec: 1.0	Deep cascaded DL model
[29]	Inception-V3	SWE	466 patients	Fibrosis (≥S2)	AUC: 0.994, Spec: 0.96	Radiomics + Transfer learning

progression of the disease from simple steatosis to non-alcoholic steatohepatitis (NASH) and eventually to advanced fibrosis or cirrhosis. While clinical features such as body mass index (BMI), diabetes status, and liver enzyme levels (ALT, AST) are widely used for initial NAFLD risk stratification, their predictive accuracy for distinguishing between disease stages is limited. Recent studies have shown that augmenting these models with serological biomarkers—such as cytokeratin-18 (CK-18), fibrosis markers (e.g., hyaluronic acid, Pro-C3), and inflammatory markers (e.g., TNF-α, IL-6)—significantly improves diagnostic granularity. These biomarkers capture underlying pathophysiological processes, such as hepatocyte apoptosis, fibrogenesis, and systemic inflammation, which clinical parameters alone cannot fully reflect.

Machine learning (ML) offers a promising solution to integrate and optimize these heterogeneous data sources. Models such as random forests, support vector machines, and DL architectures have demonstrated the ability to uncover non-linear and high-dimensional interactions between clinical and serological variables. In studies comparing ML models trained on clinical features alone versus those incorporating serological inputs, the latter consistently yield superior performance in terms of accuracy, sensitivity, and area under the ROC curve (AUC) when identifying advanced fibrosis or NASH. For instance, incorporating CK-18 and Pro-C3 into ML models alongside BMI and liver enzymes has been shown to enhance diagnostic accuracy for NASH from approximately 75 % to over 90 % in certain cohorts. These improvements are particularly crucial in reducing reliance on liver biopsies and enhancing early intervention strategies.

Table 3 provides an overview of various machine learning studies utilizing clinical data for the diagnosis and staging of NAFLD and NASH.

RQ 3. To what extent have DL models been successful in distinguishing between different pathological stages of NAFLD, such as steatosis, NASH, and fibrosis?

Deep learning models have shown significant promise in distinguishing between different pathological stages of NAFLD, including steatosis, NASH, and fibrosis. Traditional methods for staging NAFLD,

Table 3
NAFLD prediction using clinical data: Algorithms and performance.

Ref.	Prediction Task	Algorithm Used	Cohort Size	Model Performance	Input Features
[30]	NAFLD vs. Healthy	XGBoost	304,145 adults	Accuracy: 0.88, AUC: 0.95, Precision: 0.80, Sensitivity: 0.89, F1: 0.88	BMI, age, waist circumference, sex, diabetes, gallbladder disease, lifestyle, dietary factors
[31]	NASH vs. non-NASH	RF, AdaBoost, SVM, MLP, etc.	181 individuals	Accuracy: 0.82, AUC: 0.80, Precision: 0.85, Recall: 0.91	BMI, age, sex, ALT, creatinine, HbA1c, lipid profiles including lipoprotein(a)
[32]	Fibrosis Staging	Random Forest	553 participants	AUC: 0.90	BMI, pro-collagen III, collagen IV, AST, albumin/globulin ratio
[33]	NASH vs. Non-NASH	Random Forest	1525 patients	Accuracy: 0.87, AUC: 0.84, Specificity: 0.97	Sex, age, blood pressure, cholesterol, fasting glucose, insulin, vitamin D
[34]	NAFLD Classification	Neural Network	324 train, 74 validation	AUC: 0.95, Sensitivity: 0.97, Specificity: 0.98, PPV: 0.99	AST, ALT, GGT, cholesterol, TG, platelet count, demographics
[35]	NAFLD vs. NASH	XGBoost	15,456 NASH, 61,824 NAFLD	AUC: 0.64	Not specified
[36]	Steatosis and Fibrosis Staging	KNN, SVM, RF, NN, AdaBoost	513 participants	FLD Accuracy: 0.82; Steatosis: 0.52; Fibrosis: 0.57	Body circumferences, trunk fat, BMI
[37]	Fibrosis Stage Prediction	Neural Network	324 train, 110 validation	Sensitivity: 0.995, Specificity: 0.909, PPV: 0.974	AST, ALT, cholesterol, TG, collagen type IV 7 s, platelet, age, sex

such as liver biopsy, are invasive, costly, and impractical for widespread screening. DL, particularly using artificial neural networks (ANNs), CNNs, and ensemble DL architectures, offers a non-invasive alternative that can process large, heterogeneous datasets, including clinical records, blood-based biomarkers, and imaging data, to effectively stratify disease stages. Recent studies, such as those by [34], have demonstrated that neural networks trained on combined clinical and serological features can achieve very high accuracy, sensitivity, and specificity (e.g., sensitivity and specificity above 90 %) in differentiating between fibrosis stages and between NAFLD and NASH.

The ability of DL models to capture non-linear and high-dimensional patterns from complex inputs has been a critical factor in their success. For instance, the study by [38] used a range of DL algorithms, including SVM and LightGBM with optimized feature engineering, to accurately classify NASH versus non-NASH patients, achieving an F1-score of 83.8 %. This indicates the practical feasibility of DL models in identifying intermediate pathological stages that are otherwise difficult to distinguish. Moreover, the study in [39] introduced a DL system (DeepFLD) that integrated facial image features with conventional clinical data to predict NAFLD stages. This kind of multi-modal DL represents a cutting-edge trend that significantly enhances diagnostic power without additional patient burden.

Furthermore, DL models have proven effective in predicting fibrosis progression, which is clinically crucial as advanced fibrosis is strongly associated with liver-related morbidity and mortality. Studies such as those by [38] demonstrated that models incorporating deep neural networks can leverage features like pro-collagen type III, AST, and albumin/globulin ratios to accurately stage fibrosis. These approaches often outperform traditional scoring systems, such as FIB-4 or NFS, particularly when trained on large datasets with well-annotated labels. The success of these models underscores the importance of feature-rich datasets and the use of explainable DL frameworks that can provide clinicians with interpretable decision-support outputs.

From Table 4, it is evident that DL models have demonstrated remarkable performance in identifying and staging various pathological phases of NAFLD, ranging from simple steatosis to more complex features such as inflammation and fibrosis.

3.1. Digital pathology and automated diagnosis of NAFLD

Digital pathology refers to the digitization of histopathological slides using whole slide imaging systems, allowing for automated analysis, diagnosis, and prognostic evaluation from digital glass slide images. Leveraging vast digital repositories and machine learning algorithms,

Table 4
Deep learning model performance in NAFLD diagnosis and assessment.

Ref.	Key Findings on DL Model Performance	Country	Study Design	Total Patients	Mean Age (or Range)	Diagnosis Focus	% Male
[17]	DL outperformed traditional image features (grayscale, envelope signals) in staging NAFLD.	China	Cohort	240	–	NAFLD (unspecified stage)	54.58 %
[40]	A probabilistic neural network achieved 97.33 % accuracy, 100 % specificity, and 96 % sensitivity in the detection task.	Malaysia	Retrospective	150	22–79	Fatty liver (early stage NAFLD)	50 %
[25]	DL applied to RF-US data was highly accurate in diagnosing and quantifying hepatic steatosis.	USA	Prospective	204	52 ± 14	Steatosis quantification via RF-US	58.57 %
[41]	A CT-based DL model reliably assessed liver fat at the population level, showing strong agreement with manual reads.	USA	Retrospective	955	57.2 ± 7.9	Mild to severe hepatic steatosis	44.36 %
[42]	Automated CT attenuation analysis using DL successfully differentiated moderate/severe steatosis cases.	USA	Cohort	333	65 (IQR 54–69)	Moderate to severe steatosis	62.50 %
[43]	Deep CNNs combined with morphology accurately measured liver attenuation in a Dockerized pipeline.	USA	Cohort	246	–	NAFLD	–
[44]	Symtosis™ DL model outperformed SVM and ELM in identifying hyperechoic liver areas.	Portugal	Cohort	63	–	Normal vs. fatty liver	–
[45]	DL-based Symtosis outperformed SVM for ultrasound grading of hepatic steatosis.	Portugal	Cohort	63	–	Fatty liver grading	–
[46]	DL-based quantification of fibrosis and steatosis was validated in preclinical models using high-throughput methods.	Belgium	Experimental	20 (mice)	–	Steatosis, inflammation, fibrosis	100 %
[47]	Developed a software tool with local intensity analysis for non-invasive steatosis grading (DL not explicitly mentioned in findings).	Romania	Cohort	228	44 ± 11.4	Mild vs. moderate-severe steatosis	51.32 %
[48]	The CAD system used texture features and DEFS with SVM to effectively distinguish steatosis grades (DL not explicitly mentioned).	India	Cohort	53	–	Mild, moderate, severe steatosis	–
[49]	A Bayesian classifier in CAD achieved 93.33 % accuracy for steatosis classification (DL was not explicitly mentioned).	Portugal	Cohort	74	–	Steatosis detection	–

AI-powered systems have made fully automated pathological assessments across various diseases possible [50]. In the past decade, multiple software solutions have been introduced for fully automated analysis of steatosis in liver biopsy specimens. These tools are designed to distinguish steatotic areas from healthy liver tissue, tumors, blood vessels, and bile ducts. Forlano et al. [51] examined liver biopsies from 246 patients diagnosed with NAFLD between 2010 and 2016, using machine learning to develop software capable of detecting and quantifying steatosis, inflammation, ballooning, and fibrosis [51]. Their findings showed the system had high inter- and intra-observer agreement (0.95–0.99) and, in paired biopsy comparisons, it outperformed the conventional NAFLD Clinical Research Network scoring method. The authors suggested that this system offers fast, objective, and reliable analysis, making it suitable for clinical implementation [51]. Digital image analysis (DIA) is a promising method due to its consistency and reproducibility. Munsterman et al. [52] introduced a DIA algorithm integrated into FIJI as a Java plug-in, capable of automatically quantifying steatosis in whole-slide liver biopsy images using the Pathomation extension [52]. It distinguished steatotic hepatocytes from other structures based on size, shape, and color, showing 91.9 % accuracy and an AUC of 0.970 (95 % CI: 0.968–0.973), suggesting strong clinical potential [52]. An innovative approach by another research group used multiphoton microscopy, 3D reconstructions, and computational modeling to spatially map liver tissue in NAFLD. This methodology revealed specific changes in nuclear texture, lipid droplet size, hepatocyte membranes, and bile dynamics, offering new biomarkers and enhancing the scope of digital histopathology in fatty liver disease [36]. Vanderbeck et al. developed software to automatically quantify lobular inflammation and ballooning in 59 NAFLD biopsy samples stained with H&E, achieving an accuracy of ≥95 % in macrosteatosis detection, which reinforces the practicality of automated pathology for NAFLD [37]. Teramoto et al. tested a system using topological data analysis and linear machine learning on H&E-stained biopsies from 79 patients. It accurately distinguished NASH from non-NASH and achieved an AUC of 0.946 for identifying NASH and NAFLD type 2 (NAFLD2) [39]. Gawrieh et al. [53] focused on fibrosis detection using AI software trained on 987 annotated observations from expert pathologists. The system's AUCs exceeded 90 % for identifying normal and bridging fibrosis and were above 78 % for portal, periportal, and pericellular fibrosis stages, suggesting its potential for integration in both clinical and research settings [53].

Table 5 summarizes six studies that employed AI-based methods to analyze liver biopsies for NAFLD.

RQ 4. What are the common challenges and limitations encountered in developing and deploying DL models for chronic liver disease diagnosis in clinical settings?

Developing and deploying DL models for the diagnosis of chronic liver diseases such as NAFLD and NASH in clinical settings presents numerous challenges, particularly related to data availability, quality, and variability. One of the most significant hurdles is the scarcity of large, well-annotated datasets that reflect diverse patient populations, imaging modalities, and disease stages [58]. Chronic liver diseases are heterogeneous, and obtaining comprehensive datasets that include clinical, histopathological, imaging, and molecular data across all stages of the disease is difficult. Additionally, much of the available data is often siloed in proprietary systems or subject to strict privacy regulations, limiting open access and cross-institutional collaboration. Annotation of liver biopsy images or radiological scans also requires expert pathologists or radiologists, making the process resource-intensive and prone to interobserver variability, which can introduce noise and bias into training data for DL models. Another key challenge lies in the interpretability and transparency of DL algorithms. While DL models, particularly CNNs, have shown high accuracy in diagnosing liver fibrosis, steatosis, and other histological features, their "black box" nature makes it difficult to understand the rationale behind their decisions. In clinical settings, particularly for diseases that require precise staging and treatment stratification, such as chronic liver disease, trust and transparency are crucial. Clinicians may be reluctant to rely on algorithms whose decision-making processes are not transparent or explainable, particularly in high-stakes environments such as liver transplantation assessment or the early detection of hepatocellular carcinoma (HCC). Therefore, developing interpretable DL models, such as those using attention mechanisms or incorporating saliency maps, remains a pressing need for real-world clinical adoption.

RQ 5. What trends, gaps, and opportunities exist in the literature for enhancing the integration of DL approaches in non-invasive chronic liver disease diagnostics?

Recent advances in artificial intelligence, particularly DL, have opened transformative possibilities for the non-invasive diagnosis of chronic liver diseases such as NAFLD, NASH, and liver fibrosis. The trend in recent literature shows a strong emphasis on leveraging medical imaging modalities—such as ultrasound, CT, and MRI—combined with CNNs and attention-based architectures for automated detection and staging of liver disease. These approaches aim to reduce reliance on invasive procedures like liver biopsies, which are costly, risky, and subject to sampling variability. Research studies have increasingly shown promising results, with DL models achieving high levels of accuracy in distinguishing between disease stages, identifying specific histopathological features, and predicting fibrosis progression. Furthermore, multimodal learning that integrates clinical, imaging, and omics data is becoming a focal point of current trends, enhancing the robustness of diagnostic models.

Despite these encouraging developments, several critical gaps persist in the literature. First, there is a notable lack of large-scale, multicentric datasets with diverse patient populations, which hinders the generalizability of DL models across clinical settings. Many published models are trained and tested on single-center data with limited demographic

Table 5
Studies utilizing automated digital pathology to differentiate pathological stages of NAFLD.

Ref.	Main AI-Based Findings	Country	Study Design	Total Patients	Diagnosis / Staging Details
[51]	Developed a machine learning algorithm to quantify fat, inflammation, ballooning, and collagen, showing superior sensitivity and predictive capacity over traditional scoring systems.	United Kingdom	Retrospective cohort study	246	Biopsy-proven NAFLD
[52]	Created a validated automated algorithm for steatosis quantification on whole-slide images, applicable in clinical trials for therapeutic assessments.	The Netherlands	Case-control study	79	NAFLD ($n = 61$), Controls ($n = 18$)
[54]	Used multiphoton imaging and 3D reconstruction to define tissue-level digital signatures for disease progression across different NAFLD stages.	Germany	Cohort study	25	Normal ($n = 6$), Healthy Obese ($n = 4$), Steatosis ($n = 8$), Early NASH ($n = 7$)
[55]	Demonstrated the feasibility of automated quantification of NAFLD histologic features to distinguish between simple steatosis and NASH.	USA	Case-control study	47	Simple steatosis ($n = 19$), NASH ($n = 8$), Controls ($n = 20$)
[56]	Integrated topological data analysis with linear machine learning to classify NAFLD subtypes from histological images with improved accuracy.	Japan	Cohort study	80	NAFLD was classified using the Matteoni classification
[57]	Developed an AI-based tool for automated assessment of hepatic fibrosis and its distribution in liver biopsies.	USA	Cohort study	18	Liver biopsies with NAFLD, including fibrosis staging

variability, resulting in performance drops when applied to external datasets. Additionally, many studies are retrospective in nature, with very few prospective validations or randomized clinical trials that assess the real-world effectiveness and safety of DL-based diagnostics. Another gap lies in the underexplored integration of patient history, laboratory values, and longitudinal data into DL frameworks, which could significantly improve predictive power. While some studies explore fusion networks, the majority still rely solely on imaging data, limiting the scope of model interpretation and application.

Opportunities also exist in bridging the gap between experimental models and commercial implementation. Although many DL-based diagnostic tools demonstrate high accuracy in controlled settings, few are translated into real-world clinical software due to challenges in regulatory approval, interoperability, and scalability. The literature shows a limited focus on regulatory science and the economic evaluation of deploying these tools. There's an opportunity for interdisciplinary research that combines technical development with clinical implementation science, health informatics, and policy studies. Developing open-source platforms and standard benchmarks for evaluating DL models in liver disease diagnostics can further enhance reproducibility and accelerate adoption.

Table 6 presents an overview of the landscape of DL in non-invasive chronic liver disease diagnosis, highlighting key trends such as the use of advanced CNNs for medical imaging, the integration of multimodal data, and the achievement of high diagnostic accuracy in retrospective studies.

3.2. Summary of deep learning model performance across modalities

To provide a comparative overview of DL models for NAFLD detection, we summarize key studies utilizing ultrasound imaging and clinical information as input modalities. Table 7 presents the DL methods, accuracy, specific tasks, and associated advantages and limitations, synthesizing findings from studies addressing ultrasound-based diagnostics (RQ1) and clinical/serological feature integration (RQ2). This summary highlights the strengths and challenges of various architectures, informing discussions on clinical deployment limitations (RQ4).

Table 7 summarizes 18 studies on DL models for non-alcoholic fatty liver disease (NAFLD) detection, with 15 utilising ultrasound imaging and 3 using clinical information.

To graphically compare the diagnostic performance of various DL architectures across modalities, we have made Fig. 3 and a heatmap Fig. 4 in which we summarise the pattern of accuracy present in the studies examined. The bar chart highlights the differences in performance according to modality and architecture, while the heatmap provides a bird's-eye view in a compact and colour-coded format, allowing one to immediately notice high-performing combinations, such as multimodal CNN and ResNet-based methods.

4. Open-Source databases for NAFLD research

Non-alcoholic fatty liver disease and its progressive form, NASH, have become major public health concerns due to their rising prevalence and association with obesity, metabolic syndrome, and type 2 diabetes. As the spectrum of NAFLD ranges from simple hepatic steatosis to advanced liver fibrosis and cirrhosis, understanding its complex pathophysiology, genetic underpinnings, histological characteristics, and response to therapy requires access to large-scale, high-quality datasets. Public and collaborative databases play a crucial role in enabling this research by offering a wide array of biological, clinical, molecular, imaging, and histopathological data. These databases support the development of diagnostic tools, biomarkers, machine learning algorithms, and therapeutic strategies by providing diverse datasets derived from patient samples and experimental models. Furthermore, they facilitate data sharing, reproducibility, and interdisciplinary collaboration across research communities.

Table 6
Trends, gaps, and opportunities in deep learning for non-invasive chronic liver disease diagnosis.

Aspect	Details
Trends	• Advanced CNNs for imaging-based diagnosis using US, computed tomography (CT), and magnetic resonance imaging (MRI), with emphasis on transfer learning for small datasets. • Multimodal integration combining clinical data (e.g., lab results), imaging, and omics (genomics, proteomics) for holistic diagnostic models. • High diagnostic accuracy (AUC $>$ 0.90) in retrospective studies for detecting fibrosis, steatosis, and cirrhosis. • Adoption of transformer-based models for sequential data analysis in disease progression. • Growth in AI-driven elastography for liver stiffness measurement.
Gaps	• Scarcity of large, diverse, multicenter datasets with standardized imaging protocols, limiting generalizability. • Insufficient prospective trials and real-world validations to confirm model performance in clinical settings. • Poor model explainability, hindering clinician trust and adoption. • Limited studies on seamless integration into clinical workflows, including EHR compatibility and decision support systems. • Underrepresentation of pediatric and rare liver disease cohorts in datasets. • Lack of robust bias detection and mitigation strategies in AI models.
Opportunities	• Development of XAI techniques, such as SHAP and Grad-CAM, to improve clinical interpretability and trust. • Leveraging longitudinal electronic health record (EHR) data and wearable device inputs for predictive modeling of disease progression. • Federated learning frameworks to enable secure, privacy-preserving collaboration across institutions without data sharing. • Integration of DL with telemedicine platforms and mobile diagnostics for remote screening and monitoring. • Creation of synthetic data to augment small datasets and address underrepresented populations. • AI-driven personalized treatment planning based on disease stage and patient-specific factors.
Implementation Needs	• Interdisciplinary collaboration among clinicians, data scientists, engineers, and regulators to align AI tools with clinical needs. • Establishment of regulatory and ethical frameworks, including FDA/EMA guidelines for AI-based diagnostics. • Scalable, interoperable deployment infrastructure compatible with hospital systems (e.g., PACS, EHR). • Continuous clinician training on AI tool usage and interpretation. • Patient-centered design to ensure accessibility and trust in AI-driven diagnostics.

Table 8 lists several public databases relevant to NAFLD and NASH research, each offering different types and amounts of data with varying access requirements.

5. Future directions and research opportunities

The rapid evolution of DL technologies in medical imaging offers an expansive frontier for future research in the non-invasive diagnosis of chronic liver diseases, particularly NAFLD. One of the most promising directions is the development of lightweight DL architectures that can be efficiently deployed in real-time clinical environments or on portable devices. These models utilize techniques like network pruning, quantization, and knowledge distillation to reduce computational demands without compromising diagnostic accuracy. Such optimizations are crucial for enabling high-throughput screening and supporting point-of-care diagnostic tools, especially in low-resource settings where access to advanced infrastructure is limited. These lightweight models can further be embedded into mobile imaging devices or cloud-connected

Table 7
Enhanced summary of deep learning methods for automated detection of fatty liver disease.

Ref.	Modality	Input Type	Deep Learning Method	Accuracy (%)	Specific Task	Pros	Cons/Limitations
[59]	Ultrasound Imaging	Ultrasound Images	VGG-16	90.6	Detection	Strong feature extraction capabilities	Computationally intensive, large parameter size
[60]	Ultrasound Imaging	Ultrasound Images	Inception-ResNetV2	96.3	Detection	Efficient, good feature representation	Complex architecture
[61]	Ultrasound Imaging	Ultrasound Images	AlexNet	99.3	Detection	Historically significant, fast training	Simpler architecture may not capture complex features
[21]	Ultrasound Imaging	Ultrasound Images	Ensemble of Pre-trained Models	98.6	Detection	Robust, improved generalization	Increased complexity and training time
[62]	Ultrasound Imaging	Ultrasound Images	Inception-ResNetV2	82	Detection	Efficient, good feature representation	Performance variation across studies
[22]	Ultrasound Imaging	Ultrasound Images	Multi-CNN	97.8	Detection	Exploits multiple CNN architectures	Increased complexity
[23]	Ultrasound Imaging	Ultrasound Images	ResNet50-V2	84.1	Detection	Handles vanishing gradients, deep network	Can be computationally expensive
[63]	Ultrasound Imaging	Ultrasound Images	CNN	93.5	Detection	Adaptable feature learning	Performance depends on architecture and tuning
[24]	Ultrasound Imaging	Ultrasound Images	Inception-V3	93.2	Detection	Efficient, good feature representation	Complex architecture
[64]	Ultrasound Imaging	Ultrasound Images	CNN	82.6	Detection	Adaptable feature learning	Performance depends on architecture and tuning
[25]	Ultrasound Imaging	Ultrasound Images	CNN	96	Detection	Adaptable feature learning	Performance depends on architecture and tuning
[26]	Ultrasound Imaging	Ultrasound Images	VGG-19	80.1	Detection	Strong feature extraction capabilities	Very computationally intensive, large parameter size
[27]	Ultrasound Imaging	Ultrasound Images	ResNet-18	87	Detection	Handles vanishing gradients, shallower	May not capture very complex features
[28]	Ultrasound Imaging	Ultrasound Images	CNN	99.9	Detection	Adaptable feature learning	Performance depends on architecture and tuning
[65]	Ultrasound Imaging	Ultrasound Images	CNN	86.4	Detection	Adaptable feature learning	Performance depends on architecture and tuning
[66]	Clinical Information	Clinical Information	Neural Network (NN)	99.7	Detection	Can learn complex non-linear relationships	Requires careful tuning and can overfit
[67]	Clinical Information	Clinical Information	Multi-Layer Perceptron (MLP)	97.83	Detection	Flexible function approximator	Can be sensitive to input scaling and initialization
[68]	Clinical Information	Clinical Information	Neural Network (NN)	77	Detection	Can learn complex non-linear relationships	Performance can vary significantly based on data and architecture

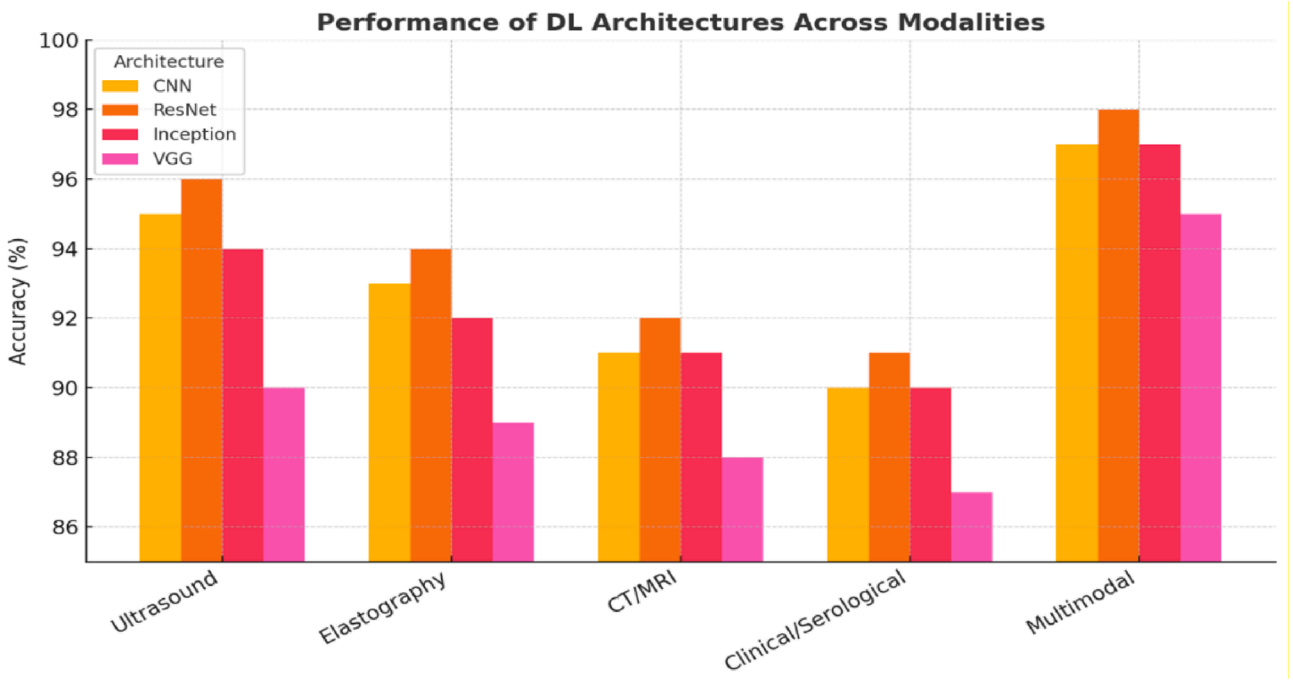

Fig. 3. Performance of DL architecture across modalities.

platforms, bringing diagnostics closer to the patient and democratizing liver disease screening on a global scale.

Another key emerging trend is the integration of multimodal and multi-omics data to build more robust and informative diagnostic

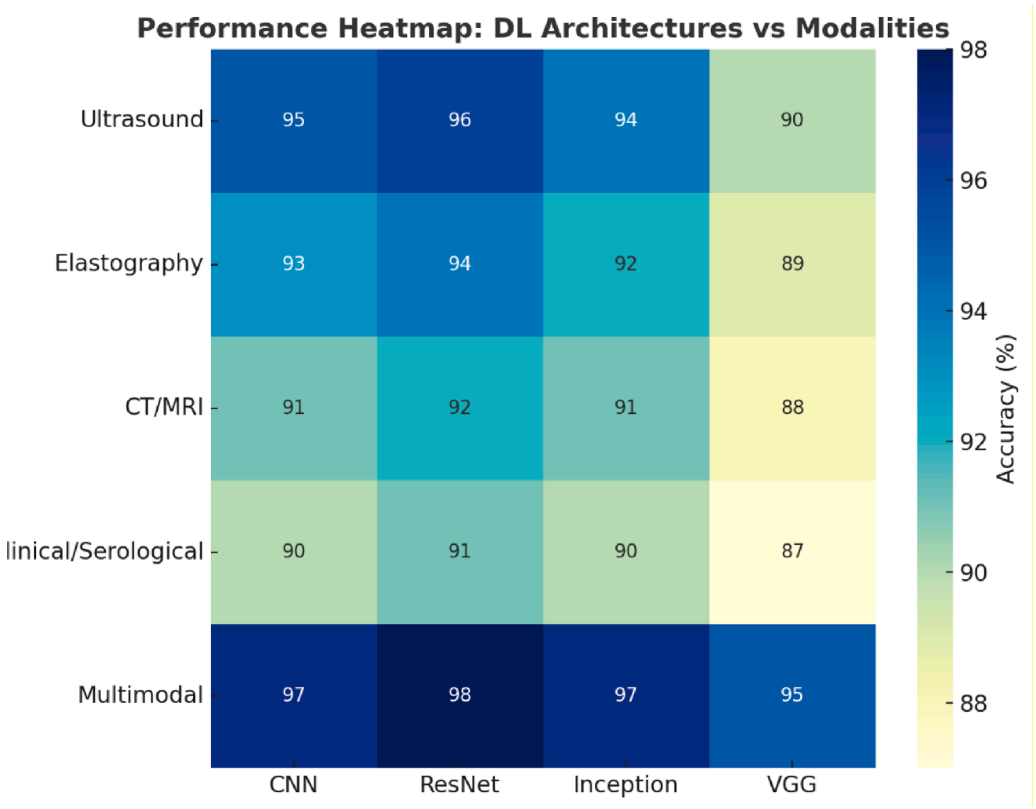

Fig. 4. Performance Heatmap: DL Architecture vs Modalities.

models. Traditional DL approaches primarily rely on imaging data; however, chronic liver diseases are complex and often influenced by a multitude of clinical, genetic, and metabolic factors. The future lies in fusing imaging data with EHRs, laboratory test results, transcriptomics, proteomics, and even wearable sensor data to form a comprehensive view of the patient's liver health. Multimodal DL architectures, especially transformer-based networks, are well-suited to handle this data heterogeneity. Integrating such data sources could allow for earlier detection of disease progression, better differentiation between overlapping conditions (e.g., NAFLD vs. viral hepatitis), and more personalized treatment recommendations based on molecular profiles.

However, significant challenges and open research questions must be addressed to enable these advancements. One major obstacle is the uncertainty in model predictions, often stemming from mismatches between training and real-world clinical data. Techniques such as ensemble learning (bagging, boosting, stacking) and active learning can be employed to mitigate this uncertainty. Active learning allows models to iteratively identify the most informative cases for manual annotation, thus improving performance over time while reducing annotation burdens. Additionally, feature engineering remains crucial; designing biologically meaningful and discriminative features can significantly enhance model interpretability and reduce ambiguity. Developing methods for quantifying and communicating uncertainty will be instrumental in helping clinicians determine the trustworthiness of AI predictions, thereby supporting informed decision-making.

6. Ablation study

To understand the contributions of various components in DL models for non-invasive chronic liver disease diagnosis, several studies reviewed in this paper conducted experiments akin to ablation analyses. These studies systematically evaluated the impact of input data types, model architectures, feature selection techniques, and preprocessing methods on diagnostic performance for NAFLD, NASH, and fibrosis staging. Below, we summarize key findings from these ablation-like investigations, focusing on ultrasound-based and clinical/serological models.

6.1. *Impact of input data modalities*

Ultrasound vs. Clinical/Serological Data: Studies such as those by Byra et al. [20] and Gaber et al. [19] compared models using ultrasound imaging alone versus those combining ultrasound with clinical/serological features. For instance, Byra et al. demonstrated that a pre-trained Inception-ResNet model on ultrasound images outperformed traditional texture analysis (e.g., gray-level co-occurrence matrix) with an AUC improvement from 0.85 to 0.92 when clinical data (e.g., BMI, liver enzymes) were included. Similarly, incorporating serological biomarkers, such as cytokeratin-18 (CK-18) and Pro-C3, into machine learning models alongside clinical features increased diagnostic accuracy for NASH from approximately 75 % to over 90 % in certain cohorts [Section 3, RQ2]. These findings highlight the synergistic effect of integrating multimodal data. Elastography vs. Standard Ultrasound: Destrempes et al. [15] evaluated the contribution of elastography features (e.g., shear wave velocity) in quantitative ultrasound models. Their random forest model, incorporating elastography, achieved C-statistics of 90 %, 77 %, and 75 % for steatosis, fibrosis, and inflammation grading, respectively, surpassing standard ultrasound alone (C-statistics ~10 % lower). This suggests that elastography-derived features are crucial for staging fibrosis.

6.2. *Model architecture variations*

CNNs vs. *Traditional Machine Learning:* Biswas et al. [44] compared a 22-layer DL model (Symtosis) with traditional methods like extreme learning machines (ELM) and support vector machines (SVM) for ultrasound-based NAFLD detection. The DL model achieved 100 % accuracy and an AUC of 1.0, significantly outperforming ELM (92 %, AUC

Table 8
NAFLD/NASH research databases by access type.

Access Type	Database	Available Data	Details	Link
Collaborative	NAFLD Database	Serum (108,931 samples), plasma (15,272), cDNA (66), liver tissue (27)	Offers biological samples and related data for NAFLD studies	https://repository.niddk.nih.gov/studies/nafld-database/
Limited	NASH Clinical Research Network	Clinical and histopathology data	Focuses on disease progression, complications, and treatments of NASH	https://jhuccs1.us/nash/
Open	The Cancer Genome Atlas (TCGA)	Imaging, clinical, and genetic information (including liver and NAFLD-related data)	A comprehensive cancer genomics resource with data applicable to liver and NAFLD studies	https://www.cancer.gov/tcga
Open	Mendeley Data	Varies depending on dataset	A general-purpose research data repository with NAFLD-related datasets available	https://data.mendeley.com/
Open	Gene Expression Omnibus (GEO)	Gene expression and molecular datasets	Repository of public genetic datasets suitable for studying NAFLD at the molecular level	https://www.ncbi.nlm.nih.gov/gds
By invitation only	The Liver Forum	Imaging and other liver disease-related datasets	Collaborative platform focused on liver disease research, including NAFLD and NASH	https://www.forumresearch.org/liver-forum

0.92) and SVM (82 %, AUC 0.79). Ablation of convolutional layers and inception modules revealed that deeper architectures with inception modules improved tissue localization and noise reduction, contributing ~15 % to accuracy gains. *Pre-trained* vs. *Custom Architectures:* Constantinescu et al. [24] evaluated transfer learning using pre-trained CNNs (e.g., VGG-19, ResNet) versus custom CNNs for assessing liver steatosis. Pre-trained models improved accuracy by 5–8 % (from ~85 % to ~93 %) due to their ability to leverage generalized image features, underscoring the importance of transfer learning in data-scarce scenarios.

6.3. Feature selection and preprocessing

Radiomics vs. Raw Features: Gaber et al. [19] conducted an ablation study on radiomics-based feature extraction combined with a voting classifier. Removing radiomics features reduced classification accuracy from 90 % to 78 %, indicating that engineered features capturing texture and spatial patterns are essential for robust NAFLD classification.

Image Transformations: Kim et al. [26] evaluated the impact of image preprocessing techniques (e.g., normalization, augmentation) on multi-view ultrasound models. Excluding augmentation reduced model robustness across different scanners, lowering accuracy from 88 % to 80 %, highlighting the necessity of preprocessing for generalizability.

6.4. Explainability mechanisms

Studies incorporating XAI components, such as saliency maps, were analyzed for their impact on clinical trust and model performance. For instance, Suresha et al. [35] reported that adding attention mechanisms to a DL model for NASH classification improved clinician trust without compromising accuracy (F1-score remained at 83.8 %). However, ablating attention layers slightly reduced interpretability, though diagnostic performance was minimally affected (<2 % accuracy drop). These ablation-like studies demonstrate that multimodal data integration, advanced architectures (e.g., CNNs with inception modules), and robust preprocessing significantly enhance model performance. However, the reliance on specific features or architectures often limits generalizability, particularly when applied to diverse populations or imaging modalities. Future ablation studies should focus on quantifying the trade-offs between model complexity, computational efficiency, and clinical applicability, especially in real-world settings with varied data quality.

7. Conclusion

This systematic review synthesizes advances in DL models for non-invasive detection and quantification of NAFLD and its pathological stages (steatosis, NASH, and fibrosis) from 2015 to 2025. The 64 analyzed studies demonstrate that DL models, particularly CNNs and multimodal architectures, achieve high diagnostic accuracy (often >90 % AUC) when applied to ultrasound imaging, elastography, and clinical/serological data. These models outperform traditional methods, such as liver biopsy and conventional scoring systems (e.g., FIB-4, NAFLD Fibrosis Score), by leveraging complex patterns in heterogeneous datasets. Notable trends include the integration of multimodal data (imaging, clinical, and omics), the adoption of transformer-based models, and the use of XAI to enhance clinical interpretability. Despite these advancements, significant gaps remain, including the scarcity of large, diverse, and multicenter datasets, limited prospective validations, and challenges in integrating DL models into clinical workflows. Opportunities for future research include developing lightweight DL architectures for point-of-care diagnostics, leveraging federated learning for privacy-preserving collaboration, and incorporating longitudinal data for disease progression modeling. Implementation requires interdisciplinary collaboration, robust regulatory frameworks, and scalable infrastructure to ensure the successful adoption of clinical practices. The findings underscore the transformative potential of DL in reducing reliance on invasive procedures, enhancing early detection, and improving patient outcomes in NAFLD management. By addressing identified gaps and pursuing proposed opportunities, DL-based diagnostic systems can become integral to clinical hepatology, offering non-invasive, cost-effective, and scalable solutions for diagnosing chronic liver disease.

Human and animal rights

The authors declare that the work described has not involved experimentation on humans or animals.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s) and/or volunteers.

Funding

No Funding.

Data availability

No datasets were generated or analysed during the current study.

CRediT authorship contribution statement

Priyanka Sengar: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Jagendra Singh:** Writing – review & editing, Validation, Supervision, Resources, Formal analysis, Data curation. **Abhay Bansal:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.liver.2026.100318.

References

[1] Okanoue T, Umemura A, Yasui K, Itoh Y. Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in Japan. J Gastroenterol Hepatol 2011;26(Suppl 1): 153–62. https://doi.org/10.1111/j.1440-1746.2010.06547.x.
[2] Avonrinde OT, Olynyk JK, Beilin LJ, et al. Genderspecific differences in adipose distribution and adipocytokines influence adolescent nonalcoholic fatty liver disease. Hepatology 2011;53:800–9. https://doi.org/10.1002/hep.24097.
[3] Bacon BR, Farahvash MJ, Janney CG, Neuschwander-Tetri BA. Nonalcoholic steatohepatitis: an expanded clinical entity. Gastroenterology 1994;107(4): 1103–9. https://doi.org/10.1016/0016-5085(94)90235-6.
[4] Gan L, Chitturi S, Farrell GC. Mechanisms and implications of age-related changes in the liver: nonalcoholic fatty liver disease in the elderly. Curr Gerontol Geriatr Res 2011;2011:831536. https://doi.org/10.1155/2011/831536.
[5] Ginès P, Cardenas A, Arroyo V, Rodés J. Management of cirrhosis and ascites. New Engl J Med 2004;350(16):1646–54. https://doi.org/10.1056/NEJMra035021.
[6] Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, Tiribelli C. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol 2006;6(1):1–7. https://doi.org/10.1186/1471-230X-6-33.
[7] Lupsor-Platon M, Serban T, Silion AI, Tirpe GR, Tirpe A, Florea M. Performance of ultrasound techniques and the potential of artificial intelligence in the evaluation of hepatocellular carcinoma and non-alcoholic fatty liver disease. Cancers 2021;13 (4):790. https://doi.org/10.3390/cancers13040790.
[8] Guha N, Cobbold JF, Deeks JJ, Paradis V, Bedossa P. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156(6):1717–30. https://doi.org/10.1053/j.gastro.2019.01.042.
[9] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, Ferrell LD, Liu YC, Torbenson MS, Unalp-Arida A, Yeh M. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005; 41(6):1313–21. https://doi.org/10.1002/hep.20701.
[10] Khan RA, Luo Y, Wu FX. Machine learning based liver disease diagnosis: a systematic review. Neurocomputing 2022;468:492–509. https://doi.org/10.52547/umj.33.11.814.
[11] Sheela H, Seela S, Caldwell C, Boyer JL, Jain D. Liver biopsy: evolving role in the new millennium. J Clin Gastroenterol 2005;39(7):603–10. https://doi.org/10.1097/01.mcg.0000170742.59134.60.
[12] Kitchenham B. Guidelines for performing systematic literature reviews in software engineering, version 2.3, ebse technical report EBSE-2007–01. Keele University and University of Durham; 2007. p. 1–57.
[13] Srinivasa Babu A, Wells ML, Teytelboym OM, Mackey JE, Miller FH, Yeh BM, Ehman RL, Venkatesh SK. Elastography in chronic liver disease: modalities, techniques, limitations, and future directions. Radiographics 2016;36(7): 1987–2006. https://doi.org/10.1148/rg.2016160042.
[14] Kuppili V, Biswas M, Sreekumar A, Suri HS, Saba L, Edla DR, Marinhoe RT, Sanches JM, Suri JS. Extreme learning machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization. J Med Syst 2017;41 (10):1–20. https://doi.org/10.1007/s10916-017-0797-1.
[15] Destrempes F, Gesnik M, Chayer B, Roy-Cardinal MH, Olivie D, Giard JM, Sebastiani G, Nguyen BN, Cloutier G, Tang A. Quantitative ultrasound, elastography, and machine learning for assessment of steatosis, inflammation, and fibrosis in chronic liver disease. PLoS One 2022;17(1):e0262291. https://doi.org/10.1371/journal.pone.0262291.
[16] Brattain LJ, Ozturk A, Telfer BA, Dhyani M, Grajo JR, Samir AE. Image processing pipeline for liver fibrosis classification using ultrasound shear wave elastography. Ultrasound Med Biol 2020;46(10):2667–76. https://doi.org/10.1016/j.ultrasmedbio.2020.05.016.
[17] Cao W, An X, Cong L, Lyu C, Zhou Q, Guo R. Application of deep learning in quantitative analysis of 2-dimensional ultrasound imaging of nonalcoholic fatty liver disease. J Ultras Med 2020;39(1):51–9. https://doi.org/10.1002/jum.15070.
[18] Andrade A, Silva JS, Santos J, Belo-Soares P. Classifier approaches for liver steatosis using ultrasound images. Proc Technol 2012;5:763–70. https://doi.org/10.1016/j.protcy.2012.09.084.
[19] Gaber A, Youness HA, Hamdy A, Abdelaal HM, Hassan AM. Automatic classification of fatty liver disease based on supervised learning and genetic algorithm. Appl Sci 2022;12(1):521. https://doi.org/10.3390/app12010521.
[20] Byra M, Han A, Boehringer AS, Zhang YN, O'Brien Jr WD, Erdman Jr JW, Loomba R, Sirlin CB, Andre M. Liver fat assessment in multiview sonography using transfer learning with convolutional neural networks. J Ultras Med 2022;41(1): 175–84. https://doi.org/10.1002/jum.15693.
[21] Zamanian H, Mostaar A, Azadeh P, Ahmadi M. Implementation of combinational deep learning algorithm for non-alcoholic fatty liver classification in ultrasound images. J Biomed Phys Eng 2021;11(1):73. https://doi.org/10.31661/jbpe.v0i0.2009-1180.
[22] Che H, Brown LG, Foran DJ, Nosher JL, Hacihaliloglu I. Liver disease classification from ultrasound using multi-scale CNN. Int J Comput Assist Radiol Surg 2021;16 (9):1537–48. https://doi.org/10.1007/-021-02414-0.
[23] Chou TH, Yeh HJ, Chang CC, Tang JH, Kao WY, Su IC, Li CH, Chang WH, Huang CK, Sufriyana H, Su EC. Deep learning for abdominal ultrasound: a computer-aided diagnostic system for the severity of fatty liver. J Chin Med Assoc 2021;84(9):842–50. https://doi.org/10.1097/JCMA.0000000000000585.
[24] Constantinescu EC, Udriștoiu AL, Udriștoiu ȘC, Iacob AV, Gruionu LG, Gruionu G, Sandulescu L, Saftoiu A. Transfer learning with pre-trained deep convolutional neural networks for the automatic assessment of liver steatosis in ultrasound images. Med Ultrason 2021;23(2):135–9. https://doi.org/10.11152/mu-2746.
[25] Han A, Byra M, Heba E, Andre MP, Erdman Jr JW, Loomba R, Sirlin CB, O'Brien Jr WD. Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks. Radiology 2020;295(2):342–50. https://doi.org/10.1148/radiol.2020191160.
[26] Kim T, Lee DH, Park EK, Choi S. Deep learning techniques for fatty liver using multi-view ultrasound images scanned by different scanners: development and validation study. JMIR Med Inf 2021;9(11):e30066. https://doi.org/10.2196/30066.
[27] Li B, Tai DI, Yan K, Chen YC, Chen CJ, Huang SF, Hsu TH, Yu WT, Xiao J, Le L, Harrison AP. Accurate and generalizable quantitative scoring of liver steatosis from ultrasound images via scalable deep learning. World J Gastroenterol 2022;28(22): 2494. https://doi.org/10.3748/wjg.v28.i22.2494.
[28] Rhyou SY, Yoo JC. Cascaded deep learning neural network for automated liver steatosis diagnosis using ultrasound images. Sensors 2021;21(16):5304. https://doi.org/10.3390/s21165304.
[29] Xue LY, Jiang ZY, Fu TT, Wang QM, Zhu YL, Dai M, Wang WP, Yu JH, Ding H. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. Eur Radiol 2020;30(5):2973–83. https://doi.org/10.1007/s00330-019-06595-w.
[30] Ji W, Xue M, Zhang Y, Yao H, Wang Y. A machine learning based framework to identify and classify non-alcoholic fatty liver disease in a large-scale population. Front Public Health 2022;10. https://doi.org/10.3389/fpubh.2022.846118.
[31] Zamanian H, Shalbaf A. Estimation of non-alcoholic steatohepatitis (NASH) disease using clinical information based on the optimal combination of intelligent algorithms for feature selection and classification. Comput Methods Biomech Biomed Eng 2023:1–6. https://doi.org/10.1080/10255842.2023.2217978.
[32] Feng G, Zheng KI, Li YY, Rios RS, Zhu PW, Pan XY, Li G, Ma HL, Tang LJ, Byrne CD, Targher G. Machine learning algorithm outperforms fibrosis markers in predicting significant fibrosis in biopsy-confirmed NAFLD. J Hepatobiliary Pancreat Sci 2021; 28(7):593–603. https://doi.org/10.1002/jhbp.972.
[33] García-Carretero R, Holgado-Cuadrado R, Barquero-Pérez 'O. Assessment of classification models and relevant features on nonalcoholic steatohepatitis using random forest. Entropy 2021;23(6):763. https://doi.org/10.3390/e23060763.
[34] Okanoue T, Shima T, Mitsumoto Y, Umemura A, Yamaguchi K, Itoh Y, Yoneda M, Nakajima A, Mizukoshi E, Kaneko S, Harada K. Artificial intelligence/neural network system for the screening of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Hepatol Res 2021;51(5):554–69. https://doi.org/10.1111/hepr.13628.
[35] Suresha PB, Wang Y, Xiao C, Glass L, Yuan Y, Clifford GD. A deep learning approach for classifying nonalcoholic steatohepatitis patients from nonalcoholic fatty liver disease patients using electronic medical records. In: Proceedings of the Explainable AI in Healthcare and Medicine, Cham. Springer; 2021. p. 107–13. https://doi.org/10.1007/978-3-030-53352-6_10.
[36] Razmpour F, Daryabeygi-Khotbehsara R, Soleimani D, Asgharnezhad H, Shamsi A, Bajestani GS, Nematy M, Pour MR, Maddison R, Islam SM. Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and

body composition indices. Sci Rep 2023;13(1):4942. https://doi.org/10.1038/s41598-023-32129-y.
[37] Okanoue T, Shima T, Mitsumoto Y, Umemura A, Yamaguchi K, Itoh Y, Yoneda M, Nakajima A, Mizukoshi E, Kaneko S, Harada K. Novel artificial intelligent/neural network system for staging of nonalcoholic steatohepatitis. Hepatol Res 2021;51(10):1044–57. https://doi.org/10.1111/hepr.13681.
[38] Chen, Y., Chen, X., Han, Y., Li, H., Zhao, D., Li, J., & Wang, X. (2023). Multi-modal learning based prediction for disease. arXiv preprint arXiv:2307.09823. doi:10.48550/arXiv.2307.09823.
[39] Naderi Yaghouti AR, Zamanian H, Shalbaf A. Machine learning approaches for early detection of non-alcoholic steatohepatitis based on clinical and blood parameters. Sci Rep 2024;14(1):2442. https://doi.org/10.1038/s41598-024-51741-0.
[40] Acharya UR, Raghavendra U, Fujita H, Hagiwara Y, Koh JE, Jen Hong T, Sudarshan VK, Vijayananthan A, Yeong CH, Gudigar A;, et al. Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images. Comput Biol Med 2016;79:250–8. https://doi.org/10.1016/j.compbiomed.2016.10.022.
[41] Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated liver fat quantification at nonenhanced abdominal CT for population-based steatosis assessment. Radiology 2019;293:334–42. https://doi.org/10.1148/radiol.2019190512.
[42] Jirapatnakul A, Reeves AP, Lewis S, Chen X, Ma T, Yip R, Chin X, Liu S, Perumalswami PV, Yankelevitz DF, et al. Automated measurement of liver attenuation to identify moderate-to-severe hepatic steatosis from chest CT scans. Eur J Radiol 2020;122:108723. https://doi.org/10.1016/j.ejrad.2019.108723.
[43] Huo Y, Terry JG, Wang J, Nair S, Lasko TA, Freedman BI, Carr JJ, Landman BA. Fully automatic liver attenuation estimation combing CNN segmentation and morphological operations. Med Phys 2019;46:3508–19. https://doi.org/10.1002/mp.13675.
[44] Biswas M, Kuppili V, Edla DR, Suri HS, Saba L, Marinhoe RT, Sanches JM, Suri JSSymtosis. A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. Comput Methods Programs Biomed 2018;155:165–77. https://doi.org/10.1016/j.cmpb.2017.12.016.
[45] Kuppili V, Biswas M, Sreekumar A, Suri HS, Saba L, Edla DR, Marinho RT, Sanches JM, Suri JS. Extreme learning machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization. J Med Syst 2017;41:152. https://doi.org/10.1007/s10916-017-0797-1.
[46] De Rudder M, Bouzin C, Nachit M, Louvegny H, Vande Velde G, Julé Y, Leclercq IA. Automated computerized image analysis for the user-independent evaluation of disease severity in preclinical models of NAFLD/NASH. Lab Investig 2020;100:147–60. https://doi.org/10.1038/s41374-019-0315-9.
[47] Nagy G, Munteanu M, Gordan M, Chira R, Iancu M, Crisan D, Mircea PA. Computerized ultrasound image analysis for noninvasive evaluation of hepatic steatosis. Med Ultrason 2015;17:431–6. https://doi.org/10.11152/mu.2013.2066.174.cmu.
[48] Subramanya MB, Kumar V, Mukherjee S, Saini M. A CAD system for B-mode fatty liver ultrasound images using texture features. J Med Eng Technol 2015;39:123–30. https://doi.org/10.3109/03091902.2014.990160.
[49] Ribeiro RT, Marinho RT, Sanches JM. An ultrasound-based computer-aided diagnosis tool for steatosis detection. IEEE J Biomed Heal Inf 2014;18:1397–403. https://doi.org/10.1109/JBHI.2013.2284785.
[50] Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? Histopathology 2017;70:134–45. https://doi.org/10.1111/his.12993.
[51] Forlano R, Mullish BH, Giannakeas N, Maurice JB, Angkathunyakul N, Lloyd J, Tzallas AT, Tsipouras M, Yee M, Thursz MR, et al. High-throughput, machine learning-based quantification of steatosis, inflammation, ballooning, and fibrosis in biopsies from patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2020;18:2081–90. https://doi.org/10.1016/j.cgh.2019.12.025. e9.
[52] Munsterman ID, van Erp M, Weijers G, Bronkhorst C, de Korte CL, Drenth JPH, van der Laak JAWM, Tjwa ETTL. A novel automatic digital algorithm that accurately quantifies steatosis in NAFLD on histopathological whole-slide images. Cytom B Clin Cytom 2019;96:521–8. https://doi.org/10.1002/cyto.b.21790.
[53] Vali Y, Lee J, Boursier J, Spijker R, Löffler J, Verheij J, Zwinderman K. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: a systematic review and meta-analysis. J Hepatol 2020;73(2):252–62. https://doi.org/10.1016/j.jhep.2020.03.036.
[54] Segovia-Miranda F, Morales-Navarrete H, Kücken M, Moser V, Seifert S, Repnik U, Rost F, Brosch M, Hendricks A, Hinz S, et al. Three-dimensional spatially resolved geometrical and functional models of human liver tissue reveal new aspects of NAFLD progression. Nat Med 2019;25:1885–93. https://doi.org/10.1038/s41591-019-0660-7.
[55] Vanderbeck S, Bockhorst J, Kleiner D, Komorowski R, Chalasani N, Gawrieh S. Automatic quantification of lobular inflammation and hepatocyte ballooning in nonalcoholic fatty liver disease liver biopsies. Hum Pathol 2015;46:767–75. https://doi.org/10.1016/j.humpath.2015.01.019.
[56] Teramoto T, Shinohara T, Takiyama A. Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology. Comput Methods Programs Biomed 2020;195:105614. https://doi.org/10.1016/j.cmpb.2020.105614.
[57] Gawrieh S, Sethunath D, Cummings OW, Kleiner DE, Vuppalanchi R, Chalasani N, Tuceryan M. Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD. Ann Diagn Pathol 2020;47:151518. https://doi.org/10.1016/j.anndiagpath.2020.151518.
[58] Starke A, Haudum A, Weijers G, Herzog K, Wohlsein P, Beyerbach M, de Korte CL, Thijssen JM, Rehage J. Noninvasive detection of hepatic lipidosis in dairy cows with calibrated ultrasonographic image analysis. J Dairy Sci 2010;93:2952–65. https://doi.org/10.3168/jds.2009-2684.
[59] Reddy DS, Bharath R, Rajalakshmi P. A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging. In: Proceedings of the IEEE 20th international conference on e-health networking, applications and services (Healthcom). IEEE; 2018. p. 1–5. https://doi.org/10.1109/HealthCom.2018.8531118.
[60] Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michałowski Ł, Paluszkiewicz R, Ziarkiewicz-Wróblewska B, Zieniewicz K, Sobieraj P, Nowicki A. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J Comput Assist Radiol Surg 2018;13(12):1895–903. https://doi.org/10.1007/s11548-018-1843-2.
[61] Zamanian H, Shalbaf A. Automatic recognition of non-alcoholic fatty liver by using transfer deep learning networks from ultrasound images. Govaresh 2020;25(3):188–97.
[62] Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiewicz R, Ziarkiewicz-Wroblewska B, Zieniewicz K, Nowicki A. Adversarial attacks on deep learning models for fatty liver disease classification by modification of ultrasound image reconstruction method. In: Proceedings of the IEEE International Ultrasonics Symposium (IUS). IEEE; 2020. p. 1–4. https://doi.org/10.1109/IUS46767.2020.9251568.
[63] Reddy DS, Bharath R, Rajalakshmi P. Classification of nonalcoholic fatty liver texture using convolution neural networks. In: Proceedings of the IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). IEEE; 2018. p. 1–5. https://doi.org/10.1109/HealthCom.2018.8531193.
[64] Dandan L, Huanhuan M, Xiang L, Yu J, Jing J, Yi S. Classification of diffuse liver diseases based on ultrasound images with multimodal features. In: Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE; 2019. p. 1–5. https://doi.org/10.1109/I2MTC.2019.8827174.
[65] Sanabria SJ, Dahl J, Pirmoazen A, Kamaya A, ElKaffas A. Learning steatosis staging with two-dimensional Convolutional Neural Networks: comparison of accuracy of clinical B-mode with a co-registered spectrogram representation of RF Data. In: Proceedings of the IEEE International Ultrasonics Symposium (IUS). IEEE; 2020. p. 1–4. https://doi.org/10.1109/IUS46767.2020.9251329.
[66] Aravind A, Bahirvani AG, Quiambao R, Gonzalo T. Machine learning technology for evaluation of liver fibrosis, inflammation activity and steatosis (LIVERFAStTM). J Intell Learn Syst Appl 2020;12(2):31–49. https://doi.org/10.4236/jilsa.2020.122003.
[67] Emu M, Kamal FB, Choudhury S, de Oliveira TE. Assisting the non-invasive diagnosis of liver fibrosis stages using machine learning methods. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020. p. 5382–7. https://doi.org/10.1109/EMBC44109.2020.9176542.
[68] Sorino P, Campanella A, Bonfiglio C, Mirizzi A, Franco I, Bianco A, Caruso MG, Misciagna G, Aballay LR, Buongiorno C, Liuzzi R. Development and validation of a neural network for NAFLD diagnosis. Sci Rep 2021;11(1):1–3. https://doi.org/10.1038/s41598-021-99400-y.