

Applications of Machine Learning in Early Identification and Prediction of Non-Alcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis

Luying Qi, Baiwang Li, Jingnan Xue, Xiangpeng Wu, Hupo Bian, Jieqiong Chen, Hongxing Zhao

Submitted to: Journal of Medical Internet Research
on: September 14, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	35

Preprint
JMIR Publications

Applications of Machine Learning in Early Identification and Prediction of Non-Alcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis

Luying Qi¹ A; Baiwang Li²; Jingnan Xue¹; Xiangpeng Wu¹; Hupo Bian¹; Jieqiong Chen¹; Hongxing Zhao¹

¹ Department of Radiology, The First Affiliated Hospital of Huzhou University Huzhou CN

² Center of Gastrointestinal Endoscopy, The Fourth People's Hospital of Jinan Jinan CN

Corresponding Author:

Hongxing Zhao

Department of Radiology, The First Affiliated Hospital of Huzhou University
No. 158, Square Back Road, Huzhou City
Huzhou
CN

Abstract

Background: Non-alcoholic fatty liver disease (NAFLD) is one of the leading causes of chronic liver disease worldwide. Research based on artificial intelligence (AI) and machine learning (ML) algorithms has increasingly been applied to the identification and diagnosis of NAFLD. Early screening for NAFLD is beneficial in reducing the progression to inflammation, liver fibrosis, and cirrhosis. To date, no study has quantitatively evaluated the application of ML algorithms in the early identification and prediction of NAFLD. Therefore, this systematic review and meta-analysis is conducted to comprehensively assess the predictive accuracy of ML in early NAFLD detection.

Objective: To systematically evaluate and quantify the performance of ML techniques in the early identification, screening, risk prediction, and diagnosis of patients with NAFLD.

Methods: A literature search was performed in the PubMed, Web of Science, Cochrane Library, and EMBASE databases from January 2020 to March 2025. The quality of studies and risk of bias were assessed via the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) and the Prediction Model Risk of Bias Assessment Tool (PROBAST). Studies reporting AUC, sensitivity, and specificity were included. Data analysis was conducted using R (version 4.3.1) and Stata (version 17.0). This study adhered to the PRISMA guidelines.

Results: A systematic review and bias assessment were performed on 23 articles, with 19 outcomes from 17 studies included in the meta-analysis. The pooled sensitivity and specificity were 0.81 [95% CI 0.76–0.85] and 0.80 [95% CI 0.76–0.84], respectively. The area under the summary receiver operating characteristic (SROC) curve (AUC) was 0.88 [95% CI 0.84–0.90]. Significant heterogeneity was observed among studies, with meta-regression indicating that reference standards and cross-validation were potential sources of heterogeneity.

Conclusions: ML demonstrated robust performance in the early identification and prediction of NAFLD, showing promise in assisting clinicians with screening and early diagnosis. However, further research is necessary due to the lack of external validation and limited exploration of certain algorithm types.

(JMIR Preprints 14/09/2025:84070)

DOI: <https://doi.org/10.2196/preprints.84070>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/preprint/84070>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/preprint/84070>



Original Manuscript

Applications of Machine Learning in Early Identification and Prediction of Non-Alcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis

Luying Qi¹, Baiwang Li², Jingnan Xue¹, Xiangpeng Wu³, Hupo Bian¹, Jieqiong Chen¹, Hongxing Zhao¹

1 Department of Radiology, The First Affiliated Hospital of Huzhou University, Huzhou 313000, Zhejiang Province, China

2 Center of Gastrointestinal Endoscopy, The Fourth People's Hospital of Jinan, Jinan 250031, Shandong Province, China

3 Department of Rehabilitation Medicine, The First Affiliated Hospital of Huzhou University, Huzhou 313000, Zhejiang Province, China

Corresponding author: Hongxing Zhao

Abstract:

Objective: To systematically evaluate and quantify the performance of ML techniques in the early identification, screening, risk prediction, and diagnosis of patients with NAFLD.

Methods: A literature search was performed in the PubMed, Web of Science, Cochrane Library, and

EMBASE databases from January 2020 to March 2025. The quality of studies and risk of bias were assessed via the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) and the Prediction Model Risk of Bias Assessment Tool (PROBAST). Studies reporting AUC, sensitivity, and specificity were included. Data analysis was conducted using R (version 4.3.1) and Stata (version 17.0). This study adhered to the PRISMA guidelines.

Results: A systematic review and bias assessment were performed on 23 articles, with 19 outcomes from 17 studies included in the meta-analysis. The pooled sensitivity and specificity were 0.81 [95% CI 0.76–0.85] and 0.80 [95% CI 0.76–0.84], respectively. The area under the summary receiver operating characteristic (SROC) curve (AUC) was 0.88 [95% CI 0.84–0.90]. Significant heterogeneity was observed among studies, with meta-regression indicating that reference standards and cross-validation were potential sources of heterogeneity.

Conclusion: ML demonstrated robust performance in the early identification and prediction of NAFLD, showing promise in assisting clinicians with screening and early diagnosis. However, further research is necessary due to the lack of external validation and limited exploration of certain algorithm types.

Keywords NAFLD AI ML Prediction Meta

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) is a major global cause of chronic liver disease, with an adult prevalence of 32%, and its incidence is projected to rise worldwide[1, 2]. The annual direct medical cost of NAFLD exceeds 100 billion dollars, posing a substantial burden on public health[3]. NAFLD encompasses a spectrum from non-alcoholic fatty liver (NAFL, simple steatosis without inflammation) to non-alcoholic steatohepatitis (NASH, coexistence of steatosis, inflammation, and hepatocyte ballooning)[4]. NASH is a risk factor for fibrosis, indicating a poor prognosis and potential progression to cirrhosis and hepatocellular carcinoma[3, 5]. Without early diagnosis and

intervention, NAFLD may advance to severe conditions such as inflammation, fibrosis, cirrhosis, and liver failure[6]. Recent nomenclature updates have replaced NAFLD with metabolic dysfunction-associated fatty liver disease (MAFLD) and metabolic dysfunction-associated steatotic liver disease (MASLD)[7]. At present, many studies still use NAFLD nomenclature. To ensure comprehensive inclusion of relevant literature, this study screened all three terms.

Liver biopsy is the gold standard for diagnosing NAFLD, but it is invasive and affected by sampling variability and inter-observer differences, making it impractical to scale to the global population of up to 1 billion affected individuals, with significant limitations in real-world applications[8]. Currently, ultrasound is recommended as a first-line screening tool for NAFLD and remains the most widely used imaging modality for diagnosing hepatic steatosis[9, 10]. While it allows subjective estimation of fat infiltration in the liver, its performance in detecting mild steatosis is suboptimal[11]. Magnetic resonance imaging (MRI) techniques can accurately quantify steatosis or identify fibrosis stages. The proton density fat fraction (PDFF) has been proposed as a standardized MR biomarker for hepatic steatosis, demonstrating high specificity for fat deposition (unconfounded by other histological or clinical parameters), but its high cost limits routine clinical use[9, 12]. FibroScan (®) enables accurate, non-invasive assessment of hepatic steatosis in NAFLD through the Controlled Attenuation Parameter (CAP), which quantifies liver fat content independently of fibrosis[11, 13]. Several serum biomarkers—such as the fatty liver index (FLI), hepatic steatosis index (HIS), and NASH Index (ION)—are also available for diagnosing or grading hepatic steatosis[10]. Although non-invasive diagnostic methods have been proposed as alternatives to liver biopsy, their clinical validity remains debated[14].

Artificial intelligence (AI) is a broad term referring to systems designed to perform advanced reasoning and analyze vast datasets with minimal human intervention[15, 16]. Machine learning (ML), a subfield of AI, focuses on developing models that learn from data to make predictions, prioritizing maximal predictive performance[17, 18]. ML enables the creation of algorithms capable

of automatically learning, identifying patterns, and generating outcome predictions[19]. While systematic reviews have explored qualitative applications of AI in NAFLD[2], quantitative studies evaluating ML algorithms for early detection, diagnosis, and prognosis of NAFLD remain lacking, with their effectiveness yet to be comprehensively assessed or synthesized. This study conducts a systematic review and meta-analysis to evaluate the accuracy and reliability of ML algorithms integrated with conventional diagnostic approaches for early NAFLD identification, aiming to provide evidence-based insights for future research.

2. Methods

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The protocol was registered on PROSPERO (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) with the report number CRD420251049731. Ethical approval was not required.

2.1 Search Strategy

To investigate the latest advancements in ML algorithms for NAFLD prediction, we conducted a comprehensive search of articles published between January 2020 and March 2025 across four databases: PubMed, Web of Science, Cochrane Library, and EMBASE. The search strategy combined theme words with free words, including "Machine Learning," "Artificial Intelligence," "Deep Learning," "Algorithms," "Data Mining," "Non-alcoholic Fatty Liver Disease," "NAFLD," "NASH," "NAFL," "MAFLD," "MASLD," and their synonyms. The detailed search strategies for all databases are provided in Supplementary Table 1.

2.2 Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) Studies evaluating NAFLD screening, early identification, risk prediction, or diagnostic accuracy; (2) Studies using ML/DL algorithms; (3) Articles evaluating algorithm performance metrics, such as sensitivity, specificity, accuracy, area under the curve (AUC), accuracy, F1 score, recall, positive predictive value (PPV), and negative predictive value

(NPV); (4) RCTs, cohort studies, case-control studies, or cross-sectional studies; (5) English-language articles; (6) Adult participants.

The exclusion criteria were as follows: (1) Reviews, animal studies, conference abstracts, case reports, editorials, protocols, commentaries, or letters; (2) Studies focusing on topics other than screening, early identification, or prediction of NAFLD (e.g., investigations of disease progression, risk stratification, severity grading, or prognosis); (3) Insufficient data for analysis (e.g., inability to extract or pool effect sizes); (4) Lack of detailed data required to reconstruct 2×2 contingency tables for diagnostic performance evaluation; (5) Outcomes involving concurrent hepatic inflammation, fibrosis, cirrhosis, or hepatocellular carcinoma (HCC).

2.3 Article Screening

All the articles were imported into EndNote X9. Duplicate records were removed through automated and manual deduplication. Screening was conducted based on titles and abstracts, followed by full-text review for eligibility. The screening process was performed independently by two researchers (Luying Qi and Jingnan Xue) and cross-checked. Discrepancies were resolved through consultation with a third researcher (Hupo Bian).

2.4 Data Extraction and Risk of Bias Assessment

From each study, the following information was extracted via predefined standardized forms: (1) Study characteristics: first author, publication year, country, study type, design, reference standard, sample size, and outcomes; (2) ML features: training set, validation set, cross-validation, external validation, ML algorithms, and algorithm performance evaluation metrics. For studies that did not report the exact values of true positives (TP), false positives (FP), false negatives (FN), or true negatives (TN), these values were manually calculated using formulas to reconstruct confusion matrices. Data extraction was conducted independently by two researchers (Baiwang Li and Xiangpeng Wu), followed by cross-checking. Discrepancies were resolved through consultation with a third researcher (Jieqiong Chen).

Study quality and risk of bias were assessed via the QUADAS-2 tool for diagnostic accuracy studies, covering four domains: patient selection, index test, reference standard, and flow&timing. Each domain was rated as high, low, or unclear risk of bias. The PROBAST tool was used to evaluate bias risk and applicability across four domains: participants, predictors, outcome, and analysis. Assessments were performed independently by two researchers (Jingnan Xue and Hupo Bian), with disagreements resolved by a third researcher (Luying Qi).

2.5 Statistical Analysis

The sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and confusion matrices were calculated for each study. Pooled results were generated via bivariate mixed-effects models to produce forest plots for diagnostic performance evaluation. Summary receiver operating characteristic (SROC) curves were plotted, and the area under the curve (AUC) with 95% confidence interval (CI) was calculated to assess overall diagnostic accuracy. Heterogeneity among studies was evaluated via the Cochran Q statistic and Higgins I² values, categorized as low, moderate, or high (25%, 50%, and 75%, respectively). Deeks' funnel plot was used to assess publication bias. Meta-regression and subgroup analyses were conducted to explore sources of heterogeneity. All analyses were performed via R (version 4.3.1) and Stata (version 17.0) software.

3. Results

3.1 Process of article screening

A total of 8,514 records were retrieved from four databases. After removing 1,895 duplicates, titles and abstracts of 6,619 articles were screened, leading to the exclusion of 6,544 records. Full texts of 75 articles were assessed for eligibility, and 23 studies[20-42] were ultimately included in the systematic review, with 17 studies[21-24, 26-29, 31, 32, 34-37, 39, 40, 42] included in the meta-analysis. The screening flowchart was presented in Figure 1.

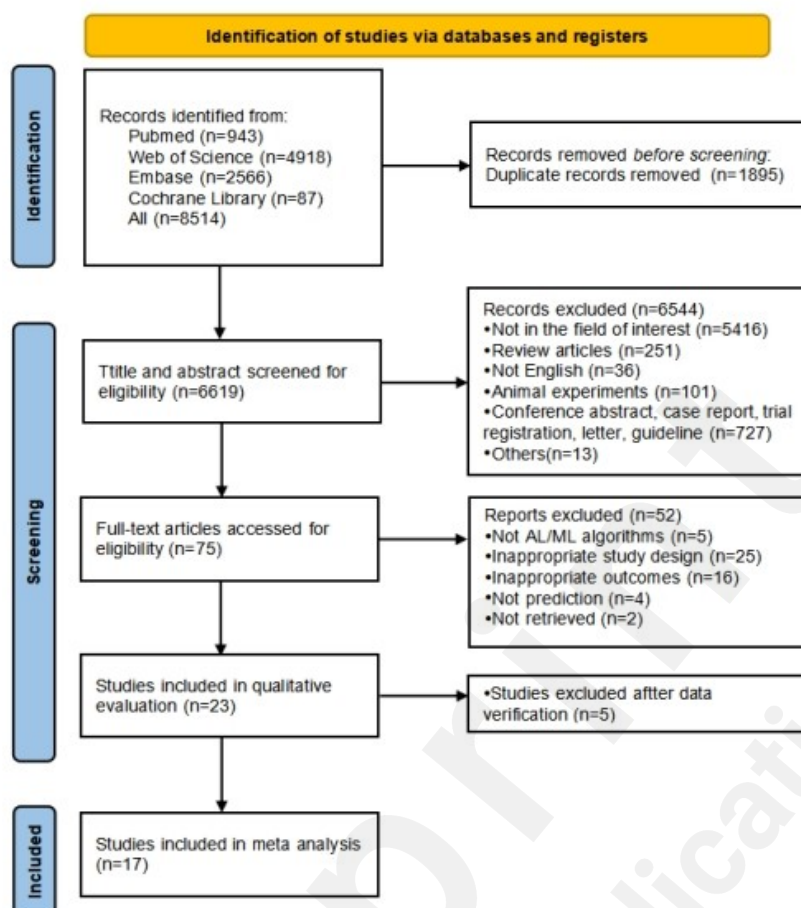


Figure 1 PRISMA Chart

3.2 Study Characteristics

Among the 23 included studies, over half were published between 2023 and 2024. The majority were conducted in China and the United States (11 in China, 6 in the US). Ten studies were cross-sectional in design. The minimum and maximum sample sizes for predictive model development were 130 and 5,171,392, respectively. Cross-validation techniques were employed in 70% of the studies, while only 30% utilized external validation datasets for model evaluation. Among all the studies, boosting algorithms were the most frequently used in optimal predictive models (6 studies). Ultrasound-based diagnostic methods accounted for 57% of the studies. Detailed characteristics of the included studies were summarized in Table 1.

Preprint
JMIR Publications

Table 1 General characteristics of included studies

First Author	Year	Country	Type	Design	Samples	Training set: Testing set	Cross validation (Y/N)	External validation (Y/N)	Option model	AUC	Reference standard	Outcome
Alizargar, A	2024	USA	Cross-sectional	Multi-center	2505	8:2	N	N	Ensemble model incorporating RF, XGBoost, and SVM	Training set: 0.981	HIS index diagnosis: HSI = 8 alanine aminotransferase/aspartate aminotransferase ratio + body mass index (BMI) + 2 if diabetes + 2 if female	NAFLD
Atsawarungruangkit, A	2021	USA	Cross-sectional	Multi-center	3235	7:3	N	N	Ensemble model of RUS boosted trees	Training set: 0.76 Testing set: 0.79	Ultrasound diagnosis	NAFLD
Tengrui Cao	2024	China	Cohort	Single-Center	22140	7:3	Y	N	XGBoost model	Training set: 0.967 Testing set: 0.914	Ultrasound diagnosis: (1)enhanced echogenicity in the near field and attenuated echogenicity in the far field of the liver parenchyma; (2)intrahepatic ducts that are relatively obscure in structure and rounded liver margins with hepatomegaly	NAFLD
Chen Chen	2024	Japan	Cross-sectional	Single-Center	4722	7:3	Y	N	RF model	Testing set: 0.889	Ultrasound diagnosis: (1)the total score of blurred blood vessels (0-1 points); (2)liver brightness (0-4 points); (3)deep attenuation (0-2 points), (4)liver and kidney echo contrast (0-4 points) exceeds 2	MASLD
Jiale Deng	2024	China	Cross-sectional	Multi-center	5171392	8:2	Y	N	CatBoost model	Training set: 0.878 Testing set: 0.862	Any of the following three criteria diagnosis: excessive overweight or obesity / type II diabetes mellitus / radiological imaging-confirmed hepatic steatosis	MAFLD
Yuhan Deng	2023	China	Cohort	Multi-center	12221	8:2	Y	Y	LSTM model	Internal validation set: 0.818	Ultrasound diagnosis	NAFLD
Fujii, T	2024	Japan	Cohort	Single-Center	186	7:3	Y	N	"clinicoradiomic" model	Training set: 0.79 Testing set: 0.80	CT diagnosis: median CT attenuation values of the liver parenchyma of < 40 Hounsfield Units (HU) without a contrast agent	NAFLD
Aiguo Han	2020	Poland	Case-control	Single-Center	204	1:1	Y	N	CNN model	Testing set: 0.98	Ultrasound RF data	NAFLD

Guoqing Huang	2023	China	Cohort	Single-Center	6196	7:3	Y	Y	LR model	Internal set: 0.778 External set: 0.806	Ultrasound diagnosis	NAFLD
Weidong Ji	2022	China	Cross-sectional	Multi-center	490980	7:3	Y	N	XGBoost model	Internal validation set: 0.951	Three criteria diagnosis: (1)subjects without drinking or drinking history; (2)no specific diseases leading to fatty liver such as viral hepatitis, liver disease induced by drug, total parenteral nutrition, hepatolenticular degeneration, and autoimmune liver disease; (3)liver imaging of subjects was consistent with the diagnostic criteria for diffuse fatty liver	NAFLD
Qingqun Li	2022	China	Case-control	Single-Center	200	7:3	Y	N	RF model	Training set: 0.945	Ultrasound diagnosis: (1)the near-field echo of the liver area is diffusely enhanced, the echo is stronger than that of a kidney; (2)the far-field echo gradually attenuates; (3)the intrahepatic duct is not displayed; (4)the liver may be slightly or moderately	NAFLD
Can Liu	2023	USA	Cross-sectional	Multi-center	6613	7:3	N	N	ANN model	Training set: 0.874 Validation set: 0.883	USFLI diagnosis: USFLI cutoff value of ≥ 30 (USFLI was calculated based on age, race, waist circumference, fasting glucose, gamma glutamyl transferase, and fasting insulin level)	NAFLD
Yuanxing Liu	2021	China	Cross-sectional	Single-Center	15315	7:3	N	N	XGBoost model	Training set: 0.926 Testing set: 0.873	Imaging or histological examination diagnosis	NAFLD
Xuefeng Ma	2021	China	Case-control	Multi-center	266	7:12	N	Y	RF model	Training set: 0.837 Validation set: 0.732	Pathology / B ultrasound / CT diagnosis	NAFLD
Masaebi, F	2024	Iran	Cohort	Multi-center	6180	7:3	Y	N	LR model	Training set: 0.92	Ultrasound diagnosis	MASLD
Hongye Peng	2023	China	Case-control	Multi-center	709	578:131	Y	Y	XGBoost model	Testing set: 0.938	Ultrasound or MRI diagnosis: ultrasound or MRI-PDFF $\geq 5\%$ according to the guideline of American Association for the Study of Liver Diseases	NAFLD
Shenghua Qin	2023	China	Cross-sectional	Single-Center	14439	7:3	Y	N	SVM model	Testing set: 0.85	Color Doppler ultrasound diagnosis	NAFLD

Sorino, P	2021	Italy	Cohort	Multi-center	2969	8:2	N	Y	NN model	Training set: 0.84 Testing set: 0.82	Ultrasound diagnosis	NAFLD
Ülger, Y	2022	Turkey	Case-control	Single-Center	400	8:2	Y	N	ANN model	Training data: 0.96 Testing data: 0.95	Clinical parameters / imaging techniques (USG/MRI) / liver biopsy diagnosis	NAFLD
Yazhi Wang	2025	USA	Cross-sectional	Multi-center	2428	7:3	Y	N	LR model	Training set: 0.856 Validation set: 0.823	CAP diagnosis: CAP ≥ 268(57–284)dB	NAFLD
Bo Yang	2024	USA	Cross-sectional	Multi-center	14913	8:2	Y	Y	LightGBM model	Training set: 0.98 Internal validation set: 0.90 External validation set: 0.81	Ultrasound diagnosis / CAP and liver stiffness measurement (LSM) diagnosis	NAFLD
Tahmasebi, A	2023	USA	Cohort	Single-Center	130	8:1:1	N	Y	AutoML model	/	Ultrasound diagnosis	NAFLD
Shang-Yu Chiang	2025	China	Case-control	Single-Center	495	6:2:2	Y	N	PBCS-ConvNeXt model	Training set: 0.88	ATI images diagnosis	NAFLD

Y:yes N:no

3.3 Quality Assessment and Publication Bias

For QUADAS-2, the results of the quality assessment were shown in Figure 2. Risks of bias and applicability concerns in the included studies primarily stemmed from patient selection, with only 2 studies rated as "low risk" across all domains and 10 studies rated as "low concern" in applicability. Details were provided in Supplementary Table S2. Using the PROBAST tool, only 3 studies were rated as "low" for both bias and applicability concerns. A summary of results was presented in Table 2, and detailed responses to all 20 signaling questions were available in Supplementary Table S3. Deeks' funnel plot indicated no significant publication bias (Figure 3).



Figure 2 The risk of bias and applicability concerns of included studies

Table 2 Summary of the risk of bias and applicability concerns for each study by PROBAST

Study(Year)	Risk of Bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of Bias	Applicability
Alizargar, A(2024)	?	-	-	-	?	+	?	-	?
Atsawarungrangkit, A(2021)	+	?	+	?	+	+	+	?	+
Tengrui Cao(2024)	+	+	+	+	+	+	+	+	+
Chen Chen(2024)	+	+	?	+	-	+	-	?	-
Jiale Deng(2024)	-	+	-	-	-	+	-	-	-
Yuhan Deng(2023)	+	+	+	+	+	+	+	+	+
Fujii, T(2024)	?	+	+	-	-	+	+	-	-
Aiguo Han(2020)	+	-	?	?	+	-	+	-	-
Guoqing Huang(2023)	+	+	+	-	+	+	+	-	+
Weidong Ji(2022)	+	+	+	?	+	+	+	□	+
Qingqun Li(2022)	+	-	-	-	+	+	+	-	+
Can Liu(2023)	+	+	-	-	+	-	-	-	-
Yuanxing Liu(2021)	+	-	?	?	+	-	?	-	-
Xuefeng Ma(2021)	+	-	-	?	+	+	-	-	-
Masaebi, F(2024)	?	+	+	?	-	+	+	?	-
Hongye Peng(2023)	+	-	+	+	+	+	+	-	+
Shenghua Qin(2023)	+	+	+	?	+	+	+	?	+
Sorino, P(2021)	?	+	?	?	+	+	?	?	?
Ülger, Y(2022)	+	-	-	-	-	+	?	-	-
Yazhi Wang(2025)	+	+	+	+	+	+	+	+	+
Bo Yang(2024)	+	+	-	+	+	+	+	-	+
Tahmasebi, A(2023)	+	-	-	?	+	-	+	-	-
Shang-Yu Chiang(2025)	?	-	?	?	?	-	+	-	-

"+" : low risk of bias/ low applicability concern; "?" : unclear risk of bias/ unclear applicability concern; "-" : high risk of bias/ high applicability concern

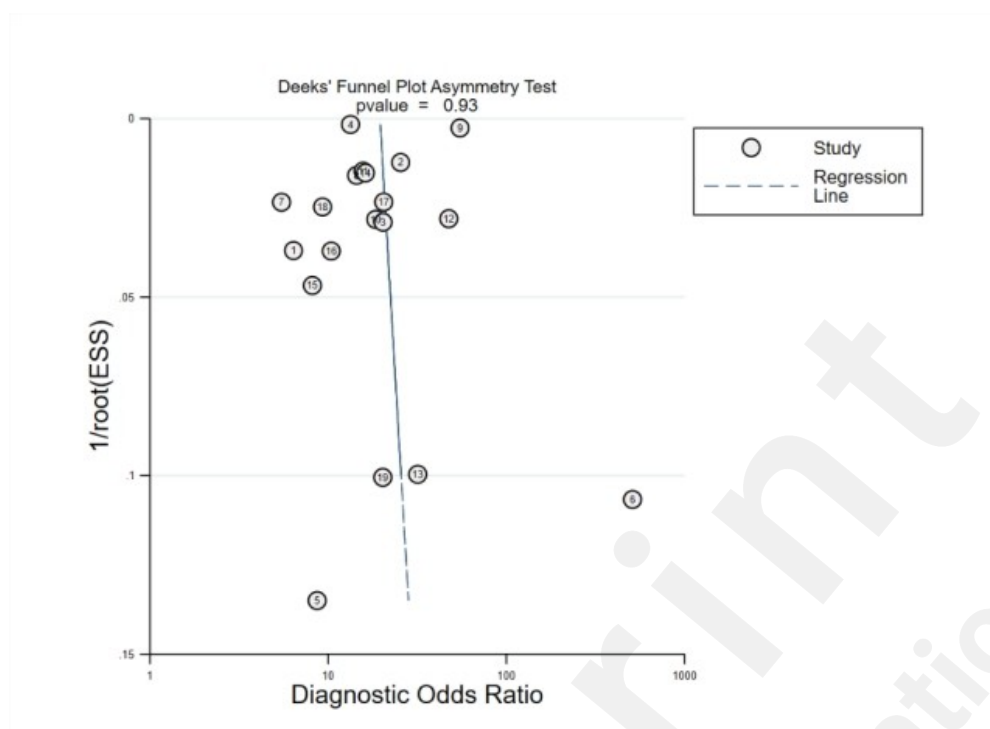


Figure 3 Deeks' funnel plot to evaluate publication bias

3.4 Predictive Performance of Meta-Analysis

After excluding studies with incomplete model evaluation metrics (AUC, TP, FP, TN, FN), the best-performing model from each study was selected based on AUC values for synthesis. For studies reporting external validation results, the models with the highest AUC values were analyzed separately. A total of 17 studies were included, with 19 outcomes subjected to meta-analysis (detailed characteristics in Table 3). The pooled sensitivity (Sen) and specificity (Spe) were 0.81 (95% CI: 0.76–0.85) and 0.80 (95% CI: 0.76–0.84), respectively, and the AUC was 0.88 (95% CI: 0.84–0.90). The bivariate forest plot was shown in Figure 4, and the SROC plot was presented in Figure 5.

Table 3 Summary of Meta analysis characteristics including internal and external validation

Author(Year)	Type of validation set	Samples	Optimal algorithm	AU C	Sensitivity	Specificity	TP	FN	TN	FP
Atsawarungruangkit, A(2021)	internal	970	Ensemble of RUS boosted trees	0.79	0.73	0.71	178	67	512	213
Tengrui Cao(2024)	internal	6641	XGBoost	0.91	0.84	0.83	2742	541	2801	557
Chen Chen(2024)	internal	1417	RF	0.89	0.82	0.82	346	78	815	178
Jiale Deng(2024)	internal	1034279	CatBoost model	0.86	0.81	0.75	84953	19412	700225	229689
Fujii, T(2024)	internal	56	"clinicoradiomic"	0.80	0.67	0.81	16	8	26	6
Aiguo Han(2020)	internal	102	CNN	0.98	0.97	0.94	68	2	30	2
Guoqing Huang(2023)	internal	1859	LR	0.78	0.71	0.70	564	236	737	322
	external	4337	LR	0.81	0.63	0.90	961	569	2513	294
Weidong Ji(2022)	internal	147294	XGBoost	0.95	0.89	0.87	65729	7804	63950	9811
Can Liu(2023)	internal	2004	ANN	0.88	0.79	0.83	1280	334	323	67
Yuanxing Liu(2021)	internal	4942	XGBoost	0.87	0.61	0.91	1136	724	2802	280
Masaebi, F(2024)	internal	1854	LR	0.92	0.86	0.89	351	57	1280	166
Hongye Peng(2023)	internal	131	XGBoost	0.94	0.92	0.77	88	9	26	8
Shenghua Qin(2023)	internal	4332	SVM	0.85	0.80	0.81	1672	431	1798	431
Sorino, P(2021)	internal	574	NN	0.82	0.65	0.82	102	56	340	76
Yazhi Wang(2025)	internal	728	LR	0.82	0.80	0.72	279	70	274	105
Bo Yang(2024)	internal	2983	LightGBM	0.90	0.93	0.61	2246	172	345	220
	external	1798	LightGBM	0.81	0.88	0.57	551	77	661	509
Shang-Yu Chiang(2025)	internal	99	PBCS-ConvNeXt	0.88	0.81	0.83	40	9	41	9

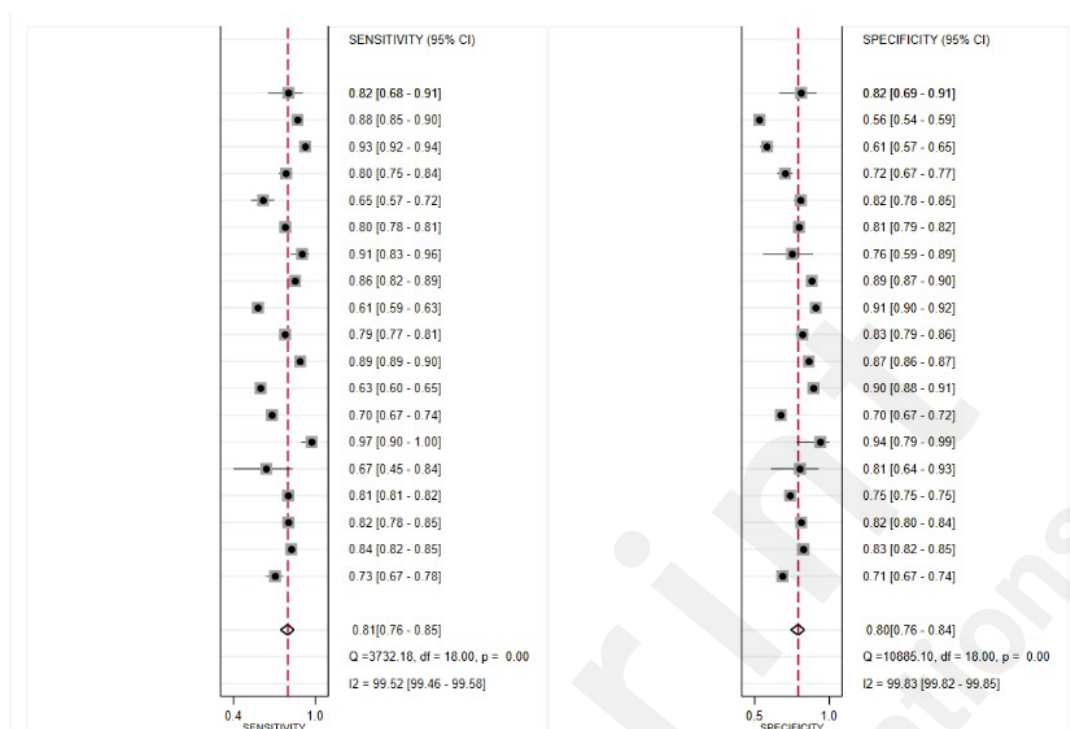


Figure 4 Coupled forest plot of pooled sensitivity and specificity

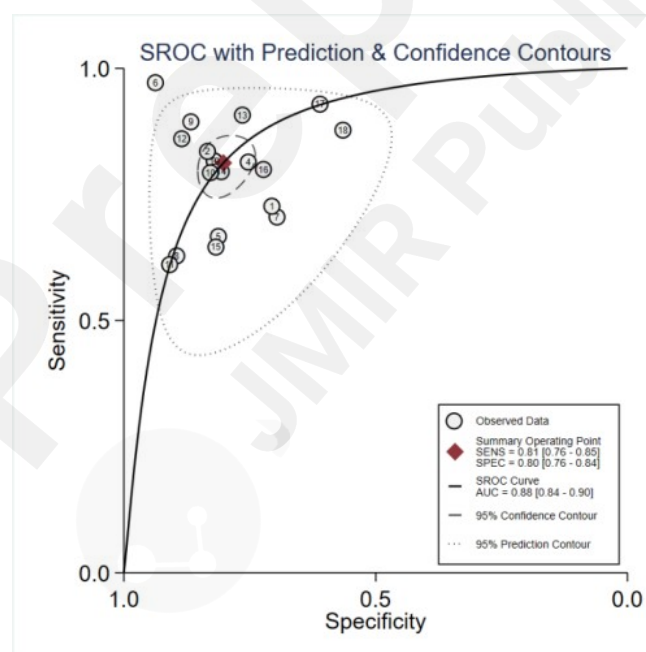


Figure 5 Summary receiver operating characteristic plots

Substantial heterogeneity was observed across studies ($I^2 > 99\%$). To explore sources of heterogeneity, meta-regression analyses were performed using reference standard and cross-validation as covariates. Results indicated that the reference standard and

cross-validation were significantly associated with heterogeneity ($P < 0.05$). The univariate meta-regression plot was shown in Figure 6. The reference standard influenced both sensitivity and specificity, with imaging-based methods showing higher sensitivity. Cross-validation also affected both sensitivity and specificity, with studies using cross-validation demonstrating higher sensitivity.



Figure 6 Univariate meta-regression plot

Subgroup analyses were conducted to assess heterogeneity across participant types, study designs, reference standards, cross-validation, validation sets, and algorithms. The results showed that general participants (AUC 0.89) exhibited slightly higher predictive accuracy compared to hospital-screened participants (AUC 0.86), potentially due to disease status variability in clinical settings. Multicenter studies (AUC 0.87) showed marginally reduced predictive accuracy, suggesting that data

collected across diverse geographic locations may introduce greater variability in outcomes. Imaging-based methods (ultrasound, CT, MRI) (AUC 0.89) demonstrated higher predictive accuracy compared to serological and other approaches (AUC 0.86). Studies without cross-validation (AUC 0.82) exhibited lower accuracy, indicating potential model overfitting and limited generalizability. The predictive accuracy of internal validation aligned with the overall estimates, likely driven by the subgroup's substantial sample size (over 80% of the total), which heavily influenced the aggregated outcomes. Models employing boosting algorithms (AUC 0.89) demonstrated robust performance, while those based on logistic regression (AUC 0.85) showed slightly inferior predictive capability. Subgroup analysis results were summarized in Table 4.

Table 4 Subgroup analysis results

Subgroup	Number s	AUC	Sensitivity	Specificity	PLR	DLR	DOR
All	19	0.88[0.84-0.90]	0.81[0.76-0.85]	0.80[0.76-0.84]	4.1[3.4-5.0]	0.23[0.19-0.30]	18[12-25]
Participants							
General	11	0.89[0.86-0.91]	0.85[0.79-0.89]	0.79[0.72-0.84]	4.0[2.9-5.3]	0.20[0.14-0.27]	20[12-34]
Hospital examination	8	0.86[0.83-0.89]	0.75[0.68-0.82]	0.83[0.77-0.87]	4.3[3.3-5.6]	0.30[0.23-0.39]	14[10-21]
Design							
Multi-center	10	0.87[0.84-0.90]	0.84[0.79-0.88]	0.77[0.70-0.82]	3.6[2.7-4.7]	0.21[0.16-0.28]	17[11-27]
Single-center	9	0.88[0.85-0.91]	0.78[0.69-0.85]	0.84[0.79-0.88]	4.9[3.7,6.6]	0.26[0.18-0.38]	19[11-32]
Reference standard							
Imaging Examination	13	0.89[0.86-0.91]	0.81[0.74-0.86]	0.83[0.79-0.86]	4.7[3.7-6.1]	0.23[0.17-0.32]	20[12-35]
Others	6	0.86[0.83-0.89]	0.82[0.74-0.89]	0.75[0.64,0.84]	3.3[2.4,4.6]	0.23[0.17,0.32]	14[11,18]
Cross validation							
Yes	15	0.89[0.86-0.91]	0.84[0.79-0.88]	0.80[0.74-0.84]	4.1[3.2-5.2]	0.20[0.16-0.27]	20[13-30]
No	4	0.82[0.79-0.85]	0.70[0.62-0.77]	0.83[0.74-0.89]	4.1[2.8-5.9]	0.36[0.29-0.45]	11[7-18]
Validation dataset							
Internal	16	0.88[0.85-0.90]	0.81[0.76-0.85]	0.81[0.77-0.85]	4.2[3.4-5.2]	0.23[0.18-0.30]	18[12-27]
Algorithm							
LR	4	0.85[0.82-0.88]	0.76[0.66-0.84]	0.82[0.71-0.89]	4.1[2.5-6.8]	0.30[0.20-0.44]	14[6-30]
Boost	7	0.89[0.86-0.91]	0.85[0.78-0.91]	0.78[0.67-0.86]	3.8[2.7-5.5]	0.19[0.13-0.27]	21[13-31]

4. Discussion

4.1 Predictive Accuracy

Our meta-analysis of 19 studies indicated that the predictive accuracy of sensitivity (0.81) and specificity (0.80) was relatively high, with the AUC of 0.88. Some previous meta-analyses using single diagnostic methods, such as Contreras et al. , discussed the accuracy of blood biomarkers in predicting NAFLD, finding that the sensitivity and specificity of the fatty liver index (FLI) were 0.67 and 0.78, respectively[43]; Khanmohammadi et al. evaluated the sensitivity and specificity of the body roundness index (BRI) in predicting NAFLD, which were 0.81 and 0.69, respectively[44]; Wang et al. reported that the sensitivity and specificity of MRI in diagnosing NAFLD were 0.82 and 0.87, respectively[45]. Decharatanachart et al. concluded that AI-assisted ultrasound had excellent performance in diagnosing NAFLD, with a sensitivity and specificity of 0.97 and 0.98, respectively[46]. The diagnostic accuracy of this AI-integrated study was significantly higher than that of the previous three studies. Our study did not directly compare the results with these earlier analyses due to methodological differences, as we integrated ML algorithms with diverse diagnostic approaches rather than focusing on a single modality.

4.2 Heterogeneity

Substantial heterogeneity was observed among the included studies. To investigate its sources, meta-regression analyses were performed on several potential factors, revealing that reference standards and cross-validation were significantly associated with heterogeneity. Although histopathological confirmation remained the gold standard for NAFLD diagnosis, its practical implementation was challenging. Consequently, many studies adopted diverse diagnostic approaches, such as imaging modalities (ultrasound, CT, MRI) and blood biomarkers, leading to considerable variability. Subgroup analyses demonstrated higher predictive accuracy for imaging-

based methods (AUC 0.89) compared to serological and other approaches (AUC 0.86), aligning with findings from Contreras, Khanmohammadi, Wang, and Decharatanachart [20-23]. These findings supported the potential utility of ML combined with imaging techniques for early NAFLD prediction. Cross-validation played a critical role in developing and evaluating ML models by mitigating overfitting[47], selecting optimal models[48], assessing generalizability[49], and facilitating hyperparameter tuning and calibration[50]. Subgroup analyses revealed lower predictive accuracy in studies without cross-validation (AUC 0.82) compared to those utilizing it (AUC 0.89), underscoring the necessity of cross-validation to enhance accuracy in ML models for early NAFLD prediction. Furthermore, models employing boosting algorithms (AUC 0.89) demonstrated promising predictive performance, consistent with studies by Bellini, Glaser, and Li et al[51-53].

4.3 Validation Sets

In this meta-analysis, the majority of studies employed internal validation methods, with only two studies utilizing external validation involving geographically independent datasets. Subgroup analyses exploring heterogeneity revealed that internal validation drove the overall pooled results, while external validation studies were insufficient in number to provide stable estimates. Since the ultimate goal of predictive models was broad applicability in general populations, the lack of external validation in these ML-based studies might limit their generalizability to real-world clinical settings for NAFLD. Thus, future research should emphasize the critical importance of external validation to ensure robustness and clinical relevance.

4.4 Clinical Value

Clinicians typically diagnosed NAFLD by synthesizing symptoms, laboratory tests, and imaging findings. For example, patients presenting with obesity, persistent

fatigue, or right upper quadrant discomfort, combined with elevated AST/ALT levels and imaging evidence of hepatic steatosis or fibrosis, might indicate NASH or fibrosis. However, early-stage NAFLD often lacked overt symptoms or significant laboratory markers of liver injury, and was frequently incidentally detected during routine screenings. While ultrasound (US) was widely used for NAFLD screening, its sensitivity for detecting mild steatosis was limited and influenced by operator-dependent variability[54]. Recent reviews highlight that AI-assisted ultrasound enhanced performance in quantifying hepatic steatosis of NAFLD patients[14], and our findings further supported that ML models integrated with imaging modalities improved diagnostic accuracy. The application of AI/ML tools could empower clinicians to make more informed and precise decisions in diagnosing early NAFLD. In the future, integrating AI/ML with imaging techniques such as ultrasound for population-wide screening could reduce underdiagnosis of early NAFLD and, through early interventions (e.g., dietary control, physical activity, and management of metabolic diseases), reverse or mitigate disease progression. This may hold significant potential to alleviate public health burdens and reduce the global prevalence of NAFLD.

5. Strengths and Limitations

To our knowledge, this systematic review and meta-analysis represented the first quantitative assessment of ML algorithms for early prediction and diagnosis of NAFLD, with several notable strengths. Firstly, given the rapid advancements in this field, we included studies published in 2020 or later to explore the latest applications of ML algorithms. Secondly, our findings indicated that ML integrated with imaging modalities provided robust predictive accuracy, supporting its potential for timely screening and preventive management. However, several limitations were identified.

Firstly, restricting inclusion to English-language studies might have excluded relevant literature. Secondly, significant heterogeneity was observed in the meta-analysis, particularly regarding reference standards and cross-validation practices. Additionally, the limited number of external validation in the included studies hindered subgroup analyses to evaluate their predictive performance.

6. Conclusion

This systematic review and meta-analysis provided the first quantitative evaluation of ML algorithms for early identification and prediction of NAFLD. The results demonstrated promising predictive accuracy of ML combined with imaging diagnostics in early NAFLD detection. ML-based tools may assist clinicians in identifying and screening early-stage NAFLD patients, enabling preventive management and timely interventions. Future research should prioritize external validation and further refinement of ML models to facilitate their translation into clinical practice.

Reference

1. Teng ML, Ng CH, Huang DQ, Chan KE, Tan DJ, Lim WH, et al. Global incidence and prevalence of nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2023;**29**:S32-s42.
2. Li Y, Wang X, Zhang J, Zhang S, Jiao J. Applications of artificial intelligence (AI) in researches on non-alcoholic fatty liver disease(NAFLD) : A systematic review. *Reviews in endocrine & metabolic disorders*. 2022;**23**:387-400.
3. Aggarwal M, Rozenbaum D, Bansal A, Garg R, Bansal P, McCullough A. Development of machine learning model to detect fibrotic non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. *Digestive and liver*

disease : *official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*. 2021;**53**:1669-72.

4. Mouskeftara T, Kalopitas G, Liapikos T, Arvanitakis K, Germanidis G, Gika H. Predicting Non-Alcoholic Steatohepatitis: A Lipidomics-Driven Machine Learning Approach. *International journal of molecular sciences*. 2024;**25**.

5. Cao C, Liu W, Guo X, Weng S, Chen Y, Luo Y, et al. Identification and validation of efferocytosis-related biomarkers for the diagnosis of metabolic dysfunction-associated steatohepatitis based on bioinformatics analysis and machine learning. *Frontiers in immunology*. 2024;**15**:1460431.

6. Alizargar A, Chang YL, Alkhaleefah M, Tan TH. Precision Non-Alcoholic Fatty Liver Disease (NAFLD) Diagnosis: Leveraging Ensemble Machine Learning and Gender Insights for Cost-Effective Detection. *Bioengineering (Basel, Switzerland)*. 2024;**11**.

7. Bilson J, Mantovani A, Byrne CD, Targher G. Steatotic liver disease, MASLD and risk of chronic kidney disease. *Diabetes & metabolism*. 2024;**50**:101506.

8. Ajmera V, Loomba R. Imaging biomarkers of NAFLD, NASH, and fibrosis. *Molecular metabolism*. 2021;**50**:101167.

9. Papatheodoridi M, Cholongitas E. Diagnosis of Non-alcoholic Fatty Liver Disease (NAFLD): Current Concepts. *Current pharmaceutical design*. 2018;**24**:4574-86.

10. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2019;**156**:1264-81.e4.

11. Ferraioli G, Soares Monteiro LB. Ultrasound-based techniques for the diagnosis of liver steatosis. *World journal of gastroenterology*. 2019;**25**:6053-62.

12. Kinner S, Reeder SB, Yokoo T. Quantitative Imaging Biomarkers of NAFLD. *Digestive diseases and sciences*. 2016;**61**:1337-47.
13. Chiu KL, Chen YD, Wang ST, Chang TH, Wu JL, Shih CM, et al. Exploring the Potential Performance of Fibroscan for Predicting and Evaluating Metabolic Syndrome using a Feature Selected Strategy of Machine Learning. *Metabolites*. 2023;**13**.
14. Alshagathrh FM, Househ MS. Artificial Intelligence for Detecting and Quantifying Fatty Liver in Ultrasound Images: A Systematic Review. *Bioengineering (Basel, Switzerland)*. 2022;**9**.
15. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism: clinical and experimental*. 2017;**69s**:S36-s40.
16. Nomura A, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial Intelligence in Current Diabetes Management and Prediction. *Current diabetes reports*. 2021;**21**:61.
17. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ (Clinical research ed)*. 2024;**385**:e078378.
18. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nature methods*. 2018;**15**:233-4.
19. Shen L, An J, Wang N, Wu J, Yao J, Gao Y. Artificial intelligence and machine learning applications in urinary tract infections identification and prediction: a systematic review and meta-analysis. *World journal of urology*. 2024;**42**:464.
20. Alizargar A, Chang YL, Alkhaleefah M, Tan TH. Precision Non-Alcoholic Fatty Liver Disease (NAFLD) Diagnosis: Leveraging Ensemble Machine Learning

and Gender Insights for Cost-Effective Detection. BIOENGINEERING-BASEL. 2024;11.

21. Atsawarungruangkit A, Laoveeravat P, Promrat K. Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database. *World journal of hepatology.* 2021;**13**:1417-27.

22. Cao T, Zhu Q, Tong C, Halengbieke A, Ni X, Tang J, et al. Establishment of a machine learning predictive model for non-alcoholic fatty liver disease: A longitudinal cohort study. *Nutrition, metabolism, and cardiovascular diseases : NMCD.* 2024;**34**:1456-66.

23. Chen C, Zhang WK, Yan GL, Tang CC. Identifying metabolic dysfunction-associated steatotic liver disease in patients with hypertension and pre-hypertension: An interpretable machine learning approach. *DIGITAL HEALTH.* 2024;**10**.

24. Deng J, Ji W, Liu H, Li L, Wang Z, Hu Y, et al. Development and validation of a machine learning-based framework for assessing metabolic-associated fatty liver disease risk. *BMC public health.* 2024;**24**:2545.

25. Deng Y, Ma Y, Fu J, Wang X, Yu C, Lv J, et al. A dynamic machine learning model for prediction of NAFLD in a health checkup population: A longitudinal study. *Heliyon.* 2023;**9**:e18758.

26. Fujii T, Iizawa Y, Kobayashi T, Hayasaki A, Ito T, Murata Y, et al. Radiomics-based prediction of nonalcoholic fatty liver disease following pancreatoduodenectomy. *Surgery today.* 2024;**54**:953-63.

27. Han A, Byra M, Heba E, Andre MP, Erdman JW, Jr., Loomba R, et al. 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**:342-50.

28. Huang G, Jin Q, Mao Y. Predicting the 5-Year Risk of Nonalcoholic Fatty Liver Disease Using Machine Learning Models: Prospective Cohort Study. *Journal of medical Internet research*. 2023;**25**:e46891.
29. 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. *Frontiers in public health*. 2022;**10**:846118.
30. Li Q, Zhang X, Zhang C, Li Y, Zhang S. Risk Factors and Prediction Models for Nonalcoholic Fatty Liver Disease Based on Random Forest. *Computational and Mathematical Methods in Medicine*. 2022;**2022**.
31. Liu C, Zhou SH, Su H, Yang WQ, Lu J. An Artificial Neural Network Model Combined with Dietary Retinol Intake from Different Sources to Predict the Risk of Nonalcoholic Fatty Liver Disease. *Biomedical and environmental sciences : BES*. 2023;**36**:1123-35.
32. Liu YX, Liu X, Cen C, Li X, Liu JM, Ming ZY, et al. Comparison and development of advanced machine learning tools to predict nonalcoholic fatty liver disease: An extended study. *Hepatobiliary & pancreatic diseases international : HBPD INT*. 2021;**20**:409-15.
33. Ma X, Yang C, Liang K, Sun B, Jin W, Chen L, et al. A predictive model for the diagnosis of non-alcoholic fatty liver disease based on an integrated machine learning method. *American Journal of Translational Research*. 2021;**13**:12704-13.
34. Masaebi F, Azizmohammad Looha M, Mohammadzadeh M, Pahlevani V, Farjam M, Zayeri F, et al. Machine-Learning Application for Predicting Metabolic Dysfunction-Associated Steatotic Liver Disease Using Laboratory and Body Composition Indicators. *Archives of Iranian medicine*. 2024;**27**:551-62.
35. Peng HY, Duan SJ, Pan L, Wang MY, Chen JL, Wang YC, et al. Development

and validation of machine learning models for nonalcoholic fatty liver disease.

Hepatobiliary & pancreatic diseases international : HBPD INT. 2023;**22**:615-21.

36. Qin S, Hou X, Wen Y, Wang C, Tan X, Tian H, et al. Machine learning classifiers for screening nonalcoholic fatty liver disease in general adults. *Scientific reports.* 2023;**13**:3638.

37. Sorino P, Campanella A, Bonfiglio C, Mirizzi A, Franco I, Bianco A, et al. Development and validation of a neural network for NAFLD diagnosis. *Scientific reports.* 2021;**11**:20240.

38. Ülger Y, Delik A. Artificial intelligence model with deep learning in nonalcoholic fatty liver disease diagnosis: genetic based artificial neural networks. *Nucleosides, nucleotides & nucleic acids.* 2023;**42**:398-406.

39. Wang Y, Wang P. Development and validation of a new diagnostic prediction model for NAFLD based on machine learning algorithms in NHANES 2017-2020.3. *Hormones (Athens, Greece).* 2025.

40. Yang B, Lu H, Ran Y. Advancing non-alcoholic fatty liver disease prediction: a comprehensive machine learning approach integrating SHAP interpretability and multi-cohort validation. *Frontiers in endocrinology.* 2024;**15**:1450317.

41. Tahmasebi A, Wang S, Wessner CE, Vu T, Liu JB, Forsberg F, et al. Ultrasound-Based Machine Learning Approach for Detection of Nonalcoholic Fatty Liver Disease. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine.* 2023;**42**:1747-56.

42. Chiang SY, Wang YW, Su PY, Chang YY, Yen HH, Chang RF. PBCS-ConvNeXt: Convolutional Network-Based Automatic Diagnosis of Non-alcoholic Fatty Liver in Abdominal Ultrasound Images. *JOURNAL OF IMAGING INFORMATICS IN MEDICINE.* 2025.

43. Contreras D, González-Rocha A, Clark P, Barquera S, Denova-Gutiérrez E. *Diagnostic accuracy of blood biomarkers and non-invasive scores for the diagnosis of NAFLD and NASH: Systematic review and meta-analysis. Annals of hepatology.* 2023;**28**:100873.
44. Khanmohammadi S, Fallahtafti P, Habibzadeh A, Ezzatollahi Tanha A, Alamdari AA, Fallahtafti P, et al. *Effectiveness of body roundness index for the prediction of nonalcoholic fatty liver disease: a systematic review and meta-analysis. Lipids in health and disease.* 2025;**24**:117.
45. Wang X, Zhang X, Ma L. *Diagnostic performance of magnetic resonance technology in detecting steatosis or fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Medicine.* 2018;**97**:e10605.
46. Decharatanachart P, Chaiteerakij R, Tiyyarattanachai T, Treeprasertsuk S. *Application of artificial intelligence in non-alcoholic fatty liver disease and liver fibrosis: a systematic review and meta-analysis. Therapeutic advances in gastroenterology.* 2021;**14**:17562848211062807.
47. Parvande S, Yeh HW, Paulus MP, McKinney BA. *Consensus features nested cross-validation. Bioinformatics (Oxford, England).* 2020;**36**:3093-8.
48. Liu Y, Liao S, Jiang S, Ding L, Lin H, Wang W. *Fast Cross-Validation for Kernel-Based Algorithms. IEEE transactions on pattern analysis and machine intelligence.* 2020;**42**:1083-96.
49. Yang X, Wang Y, Wang R, Li J. *Ensemble Feature Selection With Block-Regularized $m \times 2$ Cross-Validation. IEEE transactions on neural networks and learning systems.* 2023;**34**:6628-41.
50. Ballout N, Etievant L, Viallon V. *On the use of cross-validation for the calibration of the adaptive lasso. Biometrical journal Biometrische Zeitschrift.*

2023;**65**:e2200047.

51. Bellini V, Russo M, Domenichetti T, Panizzi M, Allai S, Bignami EG. Artificial Intelligence in Operating Room Management. *Journal of medical systems*. 2024;**48**:19.

52. Glaser K, Marino L, Stubnya JD, Bilotta F. Machine learning in the prediction and detection of new-onset atrial fibrillation in ICU: a systematic review. *Journal of anesthesia*. 2024;**38**:301-8.

53. Li X, Song Z, Chen Y, Wu J, Jiang D, Zhang Z, et al. Correction: Immune checkpoint inhibitors-related thyroid dysfunction: influencing factor analysis, prediction model development, and management strategy proposal. *Cancer immunology, immunotherapy : CII*. 2025;**74**:74.

54. Yin C, Zhang H, Du J, Zhu Y, Zhu H, Yue H. Artificial intelligence in imaging for liver disease diagnosis. *Front Med (Lausanne)*. 2025;**12**:1591523.

Supplementary Files