International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Interpretable Machine Learning Models for Early Detection of Metabolic Dysfunction–Associated Steatotic Liver Disease Using Non-Invasive Routine Clinical and Laboratory Data

Yuan-Hao You[1,2], Li-Yun Chen[1,3], Alexander Valley Chang[1,4], Yu-Shiang Lin[1]

[1]In-Service Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; [2]Department of Medical Imaging, Fu Jen Catholic University Hospital, Fu Jen Catholic University, New Taipei City, Taiwan; [3]Division of Respiratory Therapy, Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan; [4]Aurora Oral Radiology, Seattle, WA, USA

Correspondence: Yu-Shiang Lin, In-Service Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, No. 250, Wuxing Street, Xinyi District, Taipei, 110, Taiwan, Email eriklin@tmu.edu.tw

Introduction: The purpose of this study was to develop and evaluate multiple interpretable machine learning models for the early detection of metabolic dysfunction–associated steatotic liver disease (MASLD) using non-invasive routine clinical and laboratory data, with magnetic resonance imaging proton density fat fraction (MRI-PDFF) as a quantitative reference standard, and to assess the feasibility of MRI-PDFF–calibrated risk stratification as a pre-screening approach in clinical settings.

Materials and Methods: This retrospective study analyzed a de-identified cohort of 152 patients using routine clinical and laboratory data collected at Fu Jen Catholic University Hospital (FJUH), Taiwan. A total of 17 routinely available clinical and anthropometric variables were used to develop and compare across ten interpretable machine learning models, with MRI-PDFF serving as the quantitative reference standard. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1-score. Model interpretability was assessed using Shapley Additive Explanations (SHAP) at both the patient and feature levels.

Results: Logistic Regression demonstrated the highest discriminative performance (AUC = 0.873). Random Forest achieved the best overall classification performance, with an accuracy of 0.816, precision of 0.803, recall of 0.814, and an F1-score of 0.807. SHAP analysis identified the Hepatic Steatosis Index (HSI), body fat composition, age, and triglycerides (TG) as the dominant contributors to MASLD risk prediction.

Conclusion: This study demonstrates that an interpretable, MRI-PDFF–calibrated machine learning approach based on routinely available, non-invasive clinical data is feasible for MASLD risk stratification. Both transparent linear models and ensemble methods showed clinically meaningful performance, while SHAP analysis highlighted key metabolic and anthropometric factors contributing to risk. This approach may assist in prioritizing individuals for further imaging evaluation in clinical practice.

Keywords: metabolic dysfunction–associated steatotic liver disease, MASLD, magnetic resonance imaging proton density fat fraction, MRI-PDFF, machine learning, non-invasive risk stratification, explainable artificial intelligence, XAI

Introduction

Metabolic dysfunction–associated steatotic liver disease (MASLD) has rapidly emerged as a major public health problem, reflecting parallel rises in obesity, type 2 diabetes, and cardiometabolic risk worldwide.[1–3] The 2023 multi-society consensus replaced the exclusionary NAFLD/NASH terminology with MASLD/MASH to foreground metabolic risk and reduce ambiguity in clinical pathways, a change that carries implications for screening strategies and trial eligibility.[4] MASLD not only predisposes to progressive liver injury, including steatohepatitis, fibrosis, and hepatocellular carcinoma, but also markedly increases the risk of cardiovascular disease and type 2 diabetes, making it

https://doi.org/10.2147/IJGM.S604892
Received: 24 February 2026
Accepted: 3 July 2026
Published: 10 July 2026

a multisystem metabolic disorder rather than a purely hepatic condition. Early identification of MASLD is clinically valuable, as hepatic steatosis and metabolic abnormalities are largely reversible in the initial stages. Timely intervention through lifestyle modification and metabolic control can prevent progression to steatohepatitis, fibrosis, or hepatocellular carcinoma, while simultaneously reducing cardiovascular risk. In light of these considerations, clinicians require scalable, accurate, and interpretable tools to identify patients with clinically meaningful hepatic fat.

Routine imaging and blood-based indices each have limitations for first-line detection. Although liver biopsy remains the reference standard for hepatic steatosis, it is invasive and impractical for population screening. Conventional ultrasonography is inexpensive and widely available, but its performance degrades for mild fat and varies with operator and body habitus; as a result, ultrasound-based reference standards embed measurement noise into many legacy risk scores.[5] These constraints are problematic when screening at scale or comparing risk across heterogeneous scanners and sites.

In contrast, magnetic resonance imaging proton density fat fraction (MRI-PDFF) provides a quantitative, standardized estimate of liver fat that correlates with histologic steatosis grade and tracks biologically meaningful changes over time—attributes that have established it as the reference imaging biomarker in therapeutic trials.[6,7] Recent work has proposed and validated histology-anchored PDFF thresholds and examined dual "rule-out/rule-in" cut points, further supporting its use as a calibration target for non-imaging scores.[8] Yet MRI remains too costly for universal screening in most health systems, underscoring the need for MRI-PDFF–calibrated clinical tools that triage those who truly need advanced imaging.

Several indices, including the Fatty Liver Index (FLI) and Hepatic Steatosis Index (HSI), provide low-cost and imaging-independent screening for fatty liver based on routine clinical and laboratory data.[9,10] However, these scores were derived and validated against ultrasound, not a quantitative fat reference, and can misclassify patients when transported across populations or clinical settings.[5,9,10] A large-scale external validation of FLI confirmed utility but also highlighted the need for population-appropriate cutoffs and careful calibration when applied to health-check cohorts.[11]

In recent years, machine learning has been extensively applied to liver disease prediction and detection using routine clinical and imaging data. Systematic reviews and narrative overviews converge on a common pattern: most published models were trained and validated against ultrasound (US) assessments or biopsy findings, whereas substantially fewer studies calibrated models to quantitative MRI fat measurements (eg., PDFF), despite their advantages for non-invasive quantification.[12–14] Building on ultrasound and clinical data, general adult screening models using health-exam features (eg., anthropometrics and routine labs) have reported AUROC values around 0.85–0.86 with support vector machines/random forests, illustrating a strong discriminative signal available from low-cost inputs at scale.[15] Another contemporary cohort trained and validated NAFLD models using routine anthropometrics and labs, also showing strong discrimination across algorithms and validation sets and supporting scalable non-imaging screening.[16] Beyond clinical variables, imaging-based machine learning models have also demonstrated strong performance. On non-contrast abdominal CT, radiomics and deep-learning pipelines trained against quantitative CT (QCT)–derived liver fat have enabled opportunistic, workflow-embedded case finding. Broader reviews of liver imaging emphasize both improved steatosis quantification and the field's continued reliance on biopsy as the historical reference standard.[17]

Building on this foundation, more literature has successfully applied machine learning to predict fatty liver, often using routine clinical data to create models that outperform traditional scores. Some models leveraged electronic health record data to identify patients at risk of nonalcoholic steatohepatitis (NASH),[18] while several large cohort analyses have developed NAFLD risk models from routine labs and imaging features—for example, transfer-learning CNNs on ultrasound to assess NAFLD,[19–21] Outside imaging, in a cross-sectional Chinese health-check cohort of >10,000 examinees, models built from readily available features (eg., age, sex, BMI/waist measures, labs) used liver ultrasonography as the reference and demonstrated robust discrimination suitable for large-scale screening.[22] Earlier work created laboratory-parameter machine learning tools explicitly for rule-out, prioritizing parsimonious, explainable predictors and high negative predictive value so that low-risk individuals could safely defer imaging—underscoring the practical value of simple baselines in general populations.[23] Moving beyond routine laboratory tests, multi-omic models from European cohorts (IMI DIRECT) integrated genetics, proteomics, and metabolomics to outperform traditional scores; however, they were developed in non-Asian populations, highlighting portability and calibration challenges across ancestries and care settings.[24] Notably, only a few studies have focused on Asian populations. One example is a study utilizing health-system screening data in Taiwan, which similarly adopted fatty liver status derived from hospital-based screening

workflows; however, this study still adopted ultrasound as the reference standard.[25] Overall, ultrasound-based outcome definitions remain predominant in AI research on fatty liver disease. While these studies have documented ultrasound-specific advances, they also highlight inherent limitations of ultrasound-defined labeling, underscoring the need to establish models calibrated to quantitative fat references. Taken together, these observations reveal a key structural gap in label quality—most prior machine learning models are calibrated to ultrasound (operator-dependent, semi-quantitative) or to biopsy (invasive, prone to sampling variability) rather than to a reproducible, quantitative MRI fat reference such as PDFF.[12–26]

Recent studies have increasingly applied machine learning to MASLD risk identification using routine or non-invasive variables. Using data from the National Health and Nutrition Examination Survey, Zhang et al demonstrated that machine learning models can identify MASLD using large-scale population-based health data.[27] Their study incorporated multidimensional clinical, anthropometric, and laboratory variables, and feature reduction was used to derive a smaller set of routine indicators while preserving predictive information. The selected predictors reflected metabolic dysfunction, adiposity, and liver-related biochemical changes, reinforcing the biological plausibility of routine-data-based MASLD prediction. This population-level approach supports the potential application of machine learning for non-invasive screening in broader health examination settings.

Nabrdalik et al focused on patients with diabetes mellitus, a population with a particularly high MASLD burden and frequent metabolic comorbidity.[28] They developed a machine learning-assisted logistic regression model and identified a compact set of discriminative predictors, including age, BMI, diabetes-related variables, liver enzymes, platelet count, hyperuricemia, and metformin treatment. Their work supports the feasibility of using commonly collected clinical and laboratory data to stratify MASLD risk in a high-risk metabolic population. The study also highlights the potential role of interpretable, reduced-variable models for assisting risk recognition in routine diabetes care.

McTeer et al applied supervised machine learning to predict MASLD-related histological phenotypes, including MASH (metabolic dysfunction-associated steatohepatitis) and at-risk MASH, using multicenter European datasets with biopsy-derived labels.[29] Their study is notable because the diagnostic endpoints were recorded by pathologists from liver biopsy specimens, providing a rigorous reference standard for disease severity. The authors showed that routinely available clinical variables could classify clinically meaningful MASLD phenotypes. Importantly, they also found that expanded feature sets provided only limited incremental value over core clinical variables, supporting the concept that parsimonious routine-data models may be clinically useful.

Song et al extended this concept to young Asian adults, a population in whom MASLD is increasingly recognized but often underdiagnosed because liver enzymes may remain within normal ranges and imaging is not routinely performed.[30] Their externally validated model used non-invasive health check-up parameters, including BMI, blood pressure, body composition, and skeletal muscle-related indices, to predict steatotic liver disease. The study emphasizes that MASLD screening may be relevant even in apparently healthy younger populations. It also supports the broader feasibility of scalable prediction models based on simple clinical and body-composition data.

Overall, these studies indicate that machine learning models using routine or non-invasive variables can support MASLD risk identification across different clinical settings, including population-based screening, diabetes care, biopsy-characterized MASLD severity, and young Asian adults' health examinations. However, they also reveal several evident gaps in the current MASLD machine-learning literature. First, ultrasound-based outcome definitions remain predominant in MASLD machine-learning research, whereas only limited studies have calibrated predictive models against a reproducible and quantitative MRI-based fat reference such as PDFF. Although ultrasound-based studies have demonstrated the feasibility of AI-assisted MASLD prediction, ultrasound remains operator-dependent and semi-quantitative, and biopsy-based labels are limited by invasiveness and sampling variability. Second, although MASLD is highly relevant in Taiwan and East Asian populations, MRI-PDFF-calibrated machine learning studies using regional clinical data remain limited. This is important because body composition, metabolic risk profiles, and BMI-related thresholds may differ between Asian and Western populations, affecting model transportability and clinical applicability. Third, previous models have mainly relied on raw anthropometric, laboratory, imaging, or body-composition variables, while relatively few have incorporated established composite steatosis indices, such as FLI and HSI, as interpretable predictors. Collectively, these gaps motivate the present study.

This study addresses these gaps with three main contributions: (1) To our knowledge, this study represents one of the early efforts to develop a low-cost and interpretable machine learning model for MASLD calibrated to a quantitative MRI-PDFF reference standard using routinely available non-invasive clinical indicators, highlighting its potential for real-world clinical implementation. (2) The proposed machine learning model, developed using only non-invasive indicators and incorporating both the Fatty Liver Index (FLI) and Hepatic Steatosis Index (HSI) as predictors, demonstrated strong predictive performance. (3) SHAP-based explainability analysis was performed to identify key indicators influencing model predictions and to enhance interpretability in a clinical context, supporting transparent clinical decision-making.

Materials and Methods

Dataset

We conducted a retrospective model development and validation study using a de-identified cohort from Fu Jen Catholic University Hospital (FJUH) in Taiwan, comprising paired clinical/laboratory data and abdominal MRI-PDFF. The study complies with the Declaration of Helsinki and was approved by the FJUH Institutional Review Board (IRB No. FJUH114540). The dataset was analyzed as de-identified data, with informed consent waived under this approval.

Study Population and Flow

The hospital database initially contained 633 adults who received abdominal MRI with PDFF during September 2024 to June 2025. The inclusion criteria: (1) age 20–80 years; (2) availability of anthropometrics, lipid profile, liver enzymes, glycemia markers, platelets, and albumin; and (3) valid MRI-PDFF measurements. Patients were excluded based on the following criteria: (1) known chronic liver diseases other than MASLD (eg. autoimmune hepatitis); (2) excessive alcohol intake beyond MetALD criteria; (3) malignant hepatic tumor (primary or secondary) or prior liver surgery/locoregional therapy; (4) MRI-PDFF not obtained within 3 days of blood sampling and anthropometric measurement; or (5) technically inadequate MRI (motion/iron artifact not corrected by the acquisition protocol). After sequential application of all eligibility criteria, 152 patients remained for the final analysis, as shown in Figure 1.

MRI-PDFF Acquisition and Outcome Definition

MRI-PDFF was measured using a confounder-corrected multi-echo chemical shift–encoded technique on a Philips Ingenia 3.0T MRI scanner. We recorded field strength, sequence, and ensured adequate parenchymal coverage. MRI scans with motion or iron-related artifacts that were not corrected by the acquisition protocol were excluded. The primary endpoint was the presence of steatosis, defined as an MRI-PDFF ≥5.2%, reflecting histology-anchored evidence.[6,8]

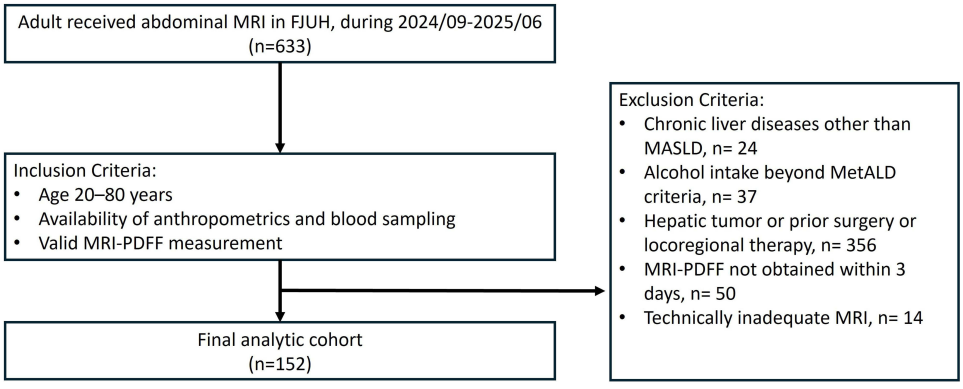

Figure 1 Flowchart illustrating the patient selection process.

Predictors and Derived Indices

Predictors were chosen for clinical availability and prior association with steatosis, including age, sex, body mass index (BMI), waist circumference, body-fat percentage, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyltransferase (GGT), triglycerides (TG), glycated hemoglobin (HbA1c), platelets, and albumin (Alb). Although FIB-4[31] and NAFLD Fibrosis Score (NFS)[32] are fibrosis staging tools rather than steatosis screening tests, they aggregate signals from age, aminotransferases, platelets, BMI, and diabetes that may carry auxiliary information about metabolic risk and disease severity; we still included them as continuous predictors to enrich the feature space. As a distinctive design setting, we also included the Fatty Liver Index (FLI) and Hepatic Steatosis Index (HSI) as training features because they are simple, imaging-independent scores built from routinely collected clinical and laboratory data. Methodologically, these indices act as high-information summary features that compress multiple correlated inputs, potentially enhancing discrimination and calibration in smaller or noisier datasets.

- Fatty Liver Index (FLI): The FLI is derived from TG (mg/dL), BMI (kg/m^2), GGT (U/L), and waist circumference (cm). Its formula is calculated as follows:

$$\mathrm{FLI} = \frac{e^{0.953\ \ln(\mathrm{TG})+0.139\,\mathrm{BMI}+0.718\ \ln(\mathrm{GGT})+0.053\,\mathrm{waist}-15.745}}{1+e^{0.953\ \ln(\mathrm{TG})+0.139\,\mathrm{BMI}+0.718\ \ln(\mathrm{GGT})+0.053\,\mathrm{waist}-15.745}} \times 100$$

- Hepatic Steatosis Index (HSI): The HSI is derived from ALT (U/L), AST (U/L), BMI (kg/m^2), sex, and diabetes status. Its formula is calculated as follows:

$$\mathrm{HSI} = 8 \times \left(\frac{\mathrm{ALT}}{\mathrm{AST}}\right) + \mathrm{BMI} + 2\,(\text{if diabetes}) + 2\,(\text{if female})$$

Model Development

This study evaluated a diverse array of ten machine learning algorithms to determine the most effective approach for MASLD classification. The selection included seven classical machine learning models and three ensemble learning architectures, aiming to provide a comprehensive assessment of performance across different algorithmic architectures. Model selection was guided by a balanced consideration of predictive accuracy and interpretability, with the goal of identifying predictive tools that combine strong performance with transparency, thereby facilitating the identification of key risk factors associated with MASLD. Five-fold cross-validation was employed to evaluate model performance. The cross-validation splits were stratified according to the MRI-PDFF-defined outcome label so that each fold maintained outcome proportions similar to those of the original dataset. In each fold, the dataset was partitioned into training and validation subsets, and all models were trained using the same cross-validation splits to ensure fair comparison. This strategy mitigates reliance on a single data split and enhances the robustness and reliability of the reported results. Given the moderate class imbalance in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training data within each fold to mitigate class imbalance during model training. The machine learning models utilized in this study are as follows.

(1) Logistic Regression
(2) Decision Tree
(3) Stochastic Gradient Descent (SGD)
(4) Support Vector Machine (SVM)
(5) Multilayer Perceptron (MLP)
(6) K-Nearest Neighbors (KNN)
(7) Artificial Neural Network (ANN)
(8) Extreme Gradient Boosting (XGBoost)
(9) Light Gradient Boosting Machine (LightGBM)

(10) Random Forest

The evaluated machine learning models were mainly implemented using default or near-default hyperparameter settings. No extensive or systematic hyperparameter optimization was performed. This strategy was adopted to maintain consistency across algorithm comparisons, improve reproducibility, and reduce the risk of overfitting or overly optimistic performance estimates due to extensive tuning, particularly given the relatively limited sample size of the present study. The main hyperparameter settings for each model are summarized in Supplementary Table S1.

Model Evaluation

Model performance was evaluated with the area under the receiver operating characteristic curve (AUROC), accuracy, precision, recall, and the F1-score. Classification metrics were derived from the confusion matrix, which details the correspondence between reference labels and model predictions. The confusion-matrix components are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The evaluation metrics were computed as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1 score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

TP refers to examinations correctly predicted as steatosis when PDFF $\geq$ 5.2%, whereas TN denotes examinations correctly predicted as non-steatosis when PDFF < 5.2%. Higher TP and TN values indicate better discrimination between steatosis and non-steatosis. Examining the confusion matrix provides practical insight into specific failure modes (FP and FN) and enables a comprehensive assessment of the model's capacity to identify hepatic steatosis in real-world imaging workflows.

Explainability

To overcome the limited interpretability of complex models and provide transparent clinical insights, we employed Shapley Additive Explanations (SHAP),[33] a widely used XAI technique. SHAP analysis was applied to the best-performing model to quantify the contribution of each input feature to the final prediction. We generated SHAP summary plots to visualize global feature importance and dependence plots to illustrate the impact of individual features on the model's output across the entire cohort.

Results

Baseline Patient Characteristics

The baseline characteristics of patients with and without MASLD are summarized in Table 1. A total of 152 patients were included in the analysis, comprising 95 patients without MASLD and 57 patients with MASLD. Categorical variables were compared using the chi-square test, while continuous variables were compared using Welch's two-sample *t*-test.

Table 1 Features of the Dataset and Baseline Characteristics of Patients

No.	Features	Absence of MASLD (n=95)	Presence of MASLD (n=57)	*p*- value
1	Gender			0.006
	Female (0)	50 (52.6%)	17 (29.8%)	
	Male (1)	45 (47.4%)	40 (70.2%)	
2	Age (years)	55.1 ± 13.5	58.0 ± 10.5	0.142
3	BMI (kg/m^2)	22.9 ± 3.2	27.1 ± 3.3	<0.001
4	Waist (cm)	81.4 ± 9.3	92.4 ± 9.1	<0.001
5	Body fat composition (%)	24.2 ± 6.1	27.9 ± 5.4	<0.001
6	AST (U/L)	22.1 ± 9.6	23.9 ± 7.3	0.191
7	ALT (U/L)	20.3 ± 14.2	29.7 ± 15.0	<0.001
8	GGT (U/L)	20.8 ± 21.2; 16.0 [12.0–21.5]	48.1 ± 103.7; 27.0 [18.0–43.0]	<0.001

(*Continued*)

Table 1 (Continued).

No.	Features	Absence of MASLD (n=95)	Presence of MASLD (n=57)	*p*- value
9	BIL-T (mg/dL)	1.0 ± 0.4	1.1 ± 0.5	0.202
10	TG (mg/dL)	101.7 ± 46.0; 94.0 [67.5–123.0]	182.3 ± 191.2; 144.0 [95.0–200.0]	<0.001
11	Platelets (10^9/L)	255.8 ± 66.0	263.7 ± 59.1	0.447
12	Albumin (g/dL)	4.5 ± 0.3	4.5 ± 0.4	0.854
13	HbA1c (%)	5.8 ± 0.6	6.0 ± 0.7	0.075
14	FLI	19.0 ± 18.7	52.2 ± 25.9	<0.001
15	FIB-4	1.2 ± 0.6	1.1 ± 0.5	0.271
16	NFS	−1.9 ± 1.5	−1.7 ± 1.3	0.388
17	HSI	30.1 ± 4.5	37.4 ± 5.3	<0.001

Abbreviations: BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyl transferase; BIL-T, total bilirubin; TG, triglycerides; HbA1c, glycated hemoglobin; FLI, Fatty Liver Index; FIB-4, Fibrosis-4 Index; NFS, NAFLD Fibrosis Score; HSI, Hepatic Steatosis Index.

Because GGT and TG showed skewed distributions, these variables were additionally summarized as median (interquartile range) and compared using the Mann–Whitney *U*-test. In this study, $p < 0.05$ was considered statistically significant. Patients with MASLD had a significantly higher proportion of males than those without MASLD (70.2% vs. 47.4%, $p = 0.006$). They also had significantly higher BMI, waist circumference, and body fat composition (all $p < 0.001$). Among laboratory parameters, ALT, GGT, and TG were significantly higher in the MASLD group. In addition, FLI and HSI were significantly higher in patients with MASLD (both $p < 0.001$). In contrast, age, total bilirubin, platelet count, albumin, HbA1c, FIB-4, and NFS did not differ significantly between groups.

Comparative Performance of Machine Learning Models

The diagnostic performance of ten distinct machine learning models was evaluated using five-fold cross-validation. The distribution of outcome labels and clinical features across the five stratified folds is presented in Supplementary Table S2. Comprehensive performance metrics, including accuracy, precision, recall, and F1-score, are summarized in Table 2. The discriminative ability of the models is illustrated by the receiver operating characteristic (ROC) curves presented in Figure 2. The curves for Logistic Regression, LightGBM, XGBoost, and Random Forest are positioned closest to the top-left corner, indicating superior discriminative performance in distinguishing between subjects with and without fatty liver. Overall, multiple models demonstrated strong discriminative performance. The Logistic Regression and LightGBM models achieved the highest AUC values (0.873 and 0.872), followed closely by XGBoost (0.861) and Random Forest (0.861). The Logistic Regression model achieved a high recall of 0.797, indicating its effectiveness in identifying true positive cases. The Random Forest model achieved the highest overall accuracy (0.816) and F1-score (0.807), suggesting

Table 2 Performance Metrics of the Evaluated Machine Learning Models for MASLD Classification

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.790	0.780	0.797	0.783
SGD	0.743	0.735	0.749	0.736
SVM	0.770	0.755	0.763	0.758
KNN	0.763	0.756	0.772	0.757
Decision Tree	0.711	0.712	0.726	0.706
Random Forest	**0.816**	**0.803**	**0.814**	**0.807**
XGBoost	0.790	0.777	0.790	0.781
LightGBM	0.763	0.750	0.761	0.754
MLP	0.737	0.725	0.737	0.728
ANN	0.743	0.726	0.721	0.723

Note: Bold values indicate the highest value for each performance metric.

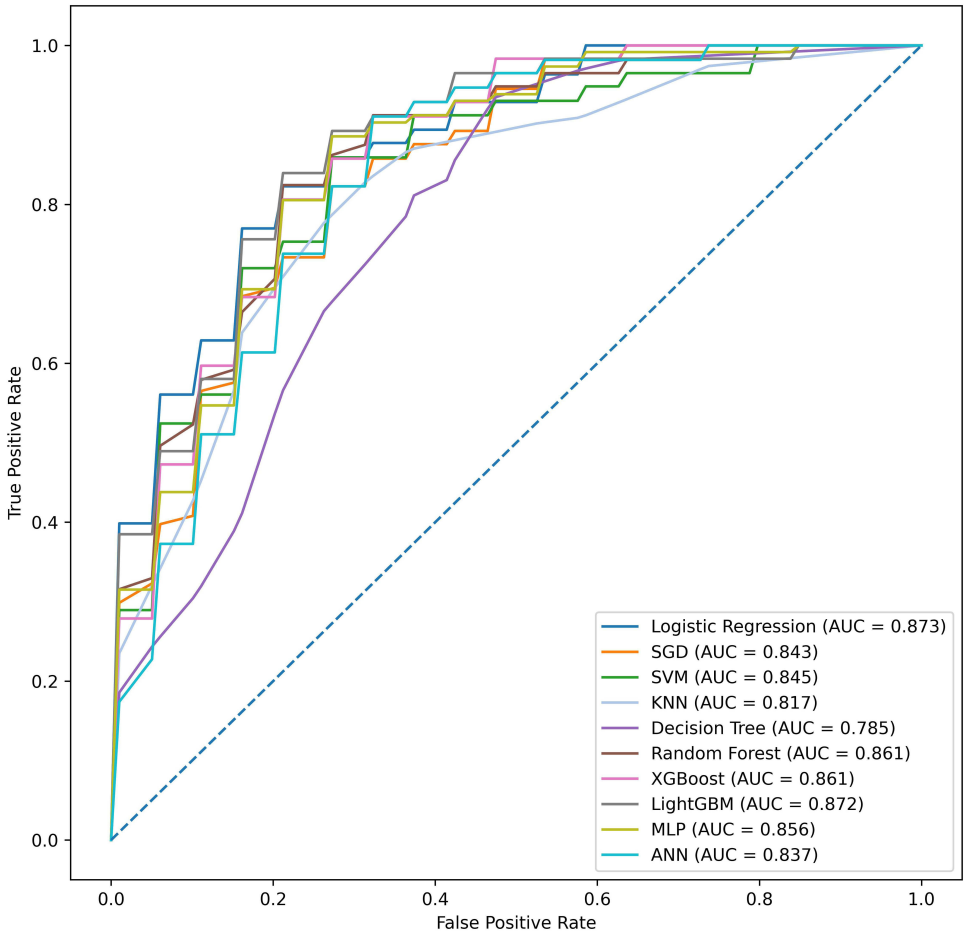

Figure 2 Receiver operating characteristic (ROC) curves of the evaluated machine learning models.

a well-balanced performance. The performance of the Support Vector Machine (SVM) was also robust, with an AUC of 0.845. In contrast, the single-tree model underperformed. The Decision Tree recorded the lowest AUC (0.785) and accuracy (0.711), with a correspondingly lower F1-score (0.706). To further assess whether the observed differences in model performance were statistically significant, fold-level AUC values from the five-fold cross-validation were compared across models using the Friedman test. The Friedman test did not show a statistically significant overall difference in AUC across models (p = 0.083). Therefore, model ranking should be interpreted descriptively based on mean AUC values rather than as evidence of statistically significant superiority.

Model Interpretability Using SHAP Analysis

To elucidate the key drivers underlying the model's predictions, SHAP analysis was performed. The SHAP summary plot (Figure 3) illustrates the global importance and impact of each feature on the model's output. The analysis revealed that the Hepatic Steatosis Index (HSI) was the most influential predictor, markedly exceeding all others and increasing predicted risk at higher values. As illustrated in the plot, higher HSI values (indicated by red dots) were associated with higher SHAP values, suggesting a higher risk of fatty liver. Following HSI, body fat composition, age, and triglycerides (TG) were the next most informative features, with higher values also associated with a higher risk of fatty liver, which aligns with clinical expectations. Notably, the FLI provided mid-tier predictive value, whereas fibrosis scores (FIB-4 and NFS) contributed only a modest auxiliary signal for steatosis classification, ranking significantly lower than HSI. Measures such as waist circumference, BMI, ALT, AST, platelets, HbA1c, and GGT showed limited average impact. Overall, the SHAP explanations support the HSI as the most informative non-invasive predictor when calibrated to the MRI-PDFF. This analysis provides a transparent view of the model's decision-making process, aligning with established clinical knowledge of MASLD risk factors.

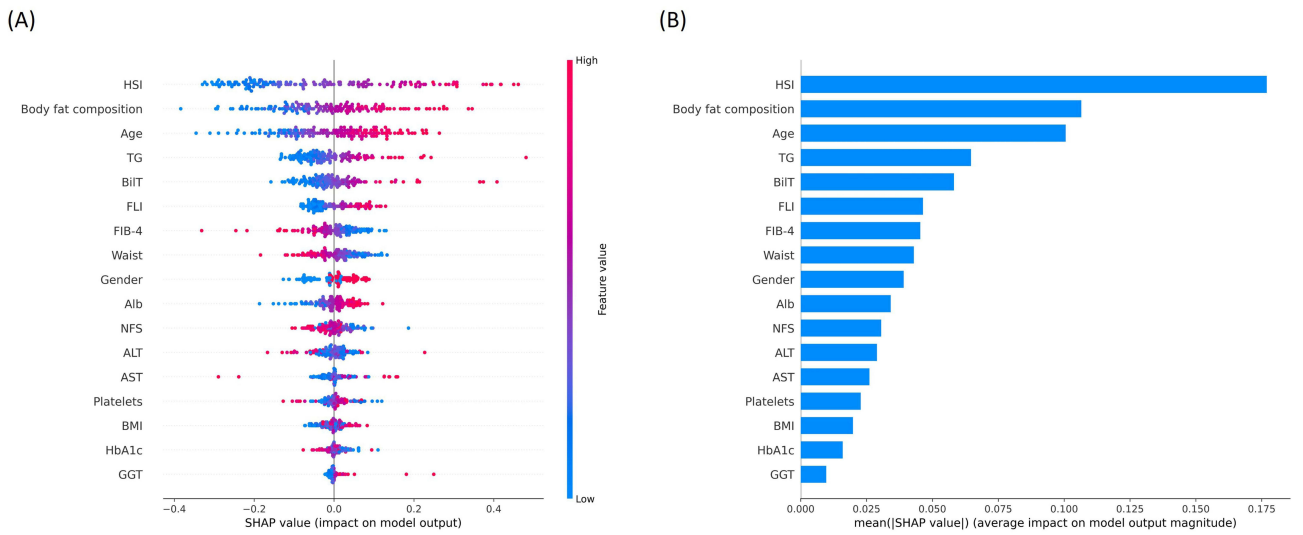

Figure 3 SHAP analysis of feature importance and model interpretation. (**A**) Beeswarm plot illustrating the distribution and impact of each feature on model predictions. (**B**) Bar plot presenting the importance ranking based on mean absolute SHAP values (|SHAP value|).

Discussion

In this study, model training and calibration were performed directly against MRI-PDFF—an imaging biomarker that quantitatively correlates with histologic steatosis—thereby better aligning the screening thresholds to a robust fat reference while maintaining clinically accessible inputs. The machine learning models demonstrated strong discrimination for hepatic steatosis. Logistic Regression yielded the highest AUC (0.873), while Random Forest demonstrated the best accuracy (0.816) and F1-score (0.807). Conversely, the Decision Tree underperformed, which may reflect the small sample size, tabular predictors, and limited spatial structure of the data.

Overall, model explanations were consistent with hepatometabolic biology. SHAP analysis ranked six features as the most influential contributors to model predictions: HSI, body fat composition, age, triglycerides (TG), BIL-T, and FLI, all of which were associated with a higher predicted risk of fatty liver. The prominence of HSI and FLI suggests both composition and population fit. Notably, HSI is distinctive in combining the ALT/AST ratio with BMI and explicitly encodes sex and diabetes status, capturing enzyme patterns and metabolic risk that are common in Asian and lean steatosis phenotypes. By contrast, FLI emphasizes adiposity and cholestatic signal, which may be less sensitive in cohorts with lower body mass yet substantial metabolic dysfunction. This interpretation is consistent with their derivation settings—HSI was developed and validated in large Korean cohorts, whereas FLI originated from an Italian general-population study.

In contrast, individual laboratory parameters such as ALT, AST, GGT, and platelets count contributed less to overall model predictions. These findings can be explained by three main factors. First, short-term biological variability may influence the results in single measurements. Second, because many metabolic markers are highly correlated, it becomes difficult to isolate the marginal effect of any single variable. Finally, there is a fundamental mismatch in what the tests measure: AST/ALT reflects active liver injury, while PDFF quantifies fat content. This distinction is important because the two are not always coupled, as many patients with hepatic steatosis have normal enzyme levels.

Two additional findings merit closer review. First, female sex was associated with a lower predicted risk of fatty liver, which is possible due to hormonal protection and lower visceral adiposity in premenopausal women. This should be investigated further by examining how the effect changes with age or menopausal status. Second, HbA1c unexpectedly showed a mildly protective effect—contrary to clinical experience. This situation may reflect post-treatment effects in these patients with diabetes, a non-linear relationship, or the information already being captured by other composite scores. Taken together, the clinical message is practical. Single laboratory markers are weak indicators for MRI-defined steatosis, but combining them into standard indices such as HSI or FLI provides a much stronger and more reliable signal. Using MRI data to guide model development ensures accuracy while relying on low-cost, routine clinical information.

https://doi.org/10.2147/IJGM.S604892

Beyond those practical messages, we also observed that, despite a relatively small sample size (n = 152) and fewer predictors, our framework performed comparably to the other larger and more elaborate models. Several mechanisms plausibly contributed to this result: the use of MRI-PDFF as a quantitative and reproducible reference reduces outcome misclassification and attenuation bias, improves the signal-to-noise ratio, and sharpens decision boundaries; the selected predictors focus hepatometabolic information that fits East Asian and lean-steatosis profiles, limiting dimensionality and multicollinearity without sacrificing clinically relevant signal; and the single-center design with harmonized acquisition and laboratory data helps constrain batch effects and distributional drift, thereby reducing irreducible noise. For future work, the sample size and population diversity could be expanded to improve model precision and stability; external, multi-vendor validation could be performed to assess generalizability across scanners, sequences, and clinical settings.

Recent studies provide an important context for interpreting the contribution of the present study. Nabrdalik et al specifically examined patients with diabetes mellitus, a clinically important high-risk subgroup for MASLD, and developed a machine learning-assisted logistic regression model using a broad set of clinical predictors.[28] Their study is valuable because it focused on a metabolically vulnerable population, used extensive baseline variables, and further demonstrated that a reduced set of discriminative predictors could still provide clinically meaningful risk stratification. In terms of study design, their study was conducted in a single European diabetes cohort, and MASLD status was defined by ultrasonographic steatosis combined with metabolic criteria. By comparison, our study used MRI-PDFF-defined hepatic steatosis as a quantitative imaging reference and evaluated routinely available anthropometric and laboratory variables in relation to directly measured liver fat. In addition, by incorporating FLI and HSI together with routine variables, our model complements the diabetes-focused approach of Nabrdalik et al and provides a screening framework calibrated to MRI-PDFF-measured hepatic fat rather than to ultrasound-defined MASLD status alone.

Zhang et al provided an important population-level reference by using NHANES data to develop machine learning models for MASLD prediction based on multidimensional but routinely obtainable clinical, anthropometric, and laboratory variables.[27] This study supports the broader concept that routinely available health-examination data contain clinically meaningful signals for MASLD screening, which is consistent with previous machine-learning studies using laboratory and clinical variables to predict fatty liver disease.[12,15,16] A major strength of Zhang et al is the use of a large, nationally representative dataset, which improves the epidemiologic relevance of the model and demonstrates the potential scalability of routine-data-based MASLD prediction. In terms of outcome definition, their outcome definition was based on transient elastography-derived controlled attenuation parameter rather than MRI-PDFF or histology. Although controlled attenuation parameter is suitable for large-scale population studies and is more standardized than conventional ultrasonography, it remains an indirect surrogate of hepatic fat and may not provide the same quantitative fat-fraction calibration as MRI-PDFF.[8,27] Notably, although the NHANES study included a much larger sample size, its best-performing model showed broadly comparable discrimination to that observed in our MRI-PDFF-calibrated study. This comparison suggests that outcome definition and alignment with quantitatively measured hepatic fat may also influence model performance, in addition to sample size.

McTeer et al further advanced the field from disease detection toward histology-based disease characterization by applying supervised learning to biopsy-derived phenotypes, including MASH and at-risk MASH, across multicenter European datasets.[29] A major strength of that study is the use of pathologist-recorded biopsy endpoints, which provide direct histological information and allow the model to address clinically important severity-related outcomes beyond simple steatosis detection. This direction is consistent with other work emphasizing machine-learning-based identification of patients at risk for progressive steatohepatitis or advanced liver disease.[18,29] Given this focus, their biopsy-based cohorts included patients with more advanced disease phenotypes, including MASH, fibrosis, or cirrhosis. By comparison, the present study addressed a different but clinically complementary task, focusing on the identification of MRI-PDFF-defined hepatic steatosis using routine anthropometric and laboratory variables before biopsy-level disease characterization is required. In addition, extended variables such as FibroScan-related measurements, autoimmune markers, or fibrosis indices are particularly informative for fibrosis or inflammatory severity, whereas the present study focused on first-line hepatic fat screening using routinely available predictors. Thus, whereas McTeer et al highlighted the potential of machine learning for staging and risk stratification across the MASLD severity spectrum, our work focuses on an earlier screening scenario using quantitative imaging calibration and interpretable low-cost predictors.

Song et al recently extended MASLD prediction to young Asian adults by developing and externally validating a machine-learning model based on non-invasive health check-up parameters, including BMI, blood pressure, percentage body fat, and skeletal muscle-related indices.[30] Their study is particularly relevant because it was conducted in an Asian population, included a large cohort, and addressed the growing need for MASLD screening in younger individuals who may have normal liver enzyme levels and may not routinely undergo imaging. Multifrequency bioelectrical impedance analysis (BIA) was used to assess body-composition markers beyond BMI and waist circumference. By comparison, their outcome was based on ultrasound-defined steatotic liver disease, whereas our study used MRI-PDFF-defined hepatic steatosis as a quantitative imaging reference. Moreover, our inclusion of FLI and HSI enabled direct evaluation of established steatosis indices alongside routine variables, providing an interpretable and easily implementable framework for first-line hepatic fat screening.

These comparisons clarify the distinct contribution of our study within the current state of MASLD machine learning research. Prior studies have shown that routine clinical, laboratory, population-level, and body-composition data can support MASLD prediction across different settings.[27–30] However, several gaps remain. First, many existing models are still trained against ultrasound- or CAP-defined steatosis rather than a quantitative fat-fraction reference. Second, biopsy-based models address clinically important but generally more advanced disease phenotypes. Third, East Asian MRI-PDFF-calibrated data remain limited. In addition, although established steatosis indices such as FLI and HSI are simple, clinically accessible, and previously validated as non-invasive screening tools, they have not been consistently incorporated into recent MASLD machine-learning pipelines together with routine clinical variables. These observations suggest that clinical usefulness depends not only on model performance but also on the reference standard, target population, predictor availability, and model interpretability.

In this context, the present study provides a complementary perspective by focusing on MRI-PDFF-defined hepatic steatosis and by evaluating whether routinely available non-invasive variables, together with composite indicators such as FLI and HSI, can provide clinically useful predictive information in a Taiwanese population. This distinction is clinically relevant because MRI-PDFF provides a quantitative assessment of hepatic fat content and may reduce outcome-label uncertainty compared with more operator-dependent or surrogate-based definitions of steatosis. At the same time, predictors derived from routine anthropometric and biochemical measurements may improve transparency, accessibility, and practical feasibility for first-line screening. Accordingly, our model may serve as a low-cost and interpretable screening tool to identify individuals who may require advanced imaging or comprehensive metabolic evaluation, particularly in settings where MRI-PDFF is accurate but not feasible as a universal screening tool.

To our knowledge, this study represents one of the early efforts to develop a low-cost and interpretable machine learning model for metabolic dysfunction–associated steatotic liver disease (MASLD), calibrated against the quantitative MRI-PDFF reference standard and based solely on routinely available, non-invasive clinical indicators. The proposed model, which integrates the Fatty Liver Index (FLI) and Hepatic Steatosis Index (HSI) derived from non-invasive indicators, demonstrated strong predictive performance (AUC = 0.873). Furthermore, SHAP-based explainability analysis was applied to identify key indicators influencing the predictions, thereby enhancing the model's clinical interpretability and credibility. Collectively, our findings suggest that MRI-PDFF-calibrated, composite indicator-based machine learning may serve as a practical first-line screening framework for identifying individuals who may benefit from further imaging-based assessment or closer metabolic risk evaluation.

Conclusion

In this study, we developed a low-cost and interpretable machine learning framework for metabolic dysfunction-associated steatotic liver disease (MASLD), calibrated against the quantitative MRI-PDFF reference standard and based solely on routinely available, non-invasive clinical indicators. The proposed framework demonstrated strong discriminative performance and clinically coherent explainability patterns, with SHAP analysis identifying established steatosis-related indicators, including HSI, body fat composition, age, triglycerides (TG), total bilirubin (BIL-T), and FLI, as key contributors. The originality of this study lies in integrating routine clinical variables with composite steatosis indices and calibrating the model against MRI-PDFF-measured hepatic fat rather than ultrasound-defined labels. These findings support the potential role of MRI-PDFF-calibrated, composite indicator-based machine learning as a practical first-line screening approach for identifying individuals who may benefit from further imaging-based assessment or closer metabolic risk evaluation.

Data Sharing Statement

The data underlying this study are not publicly available due to ethical, privacy, and institutional data governance restrictions. An anonymized dataset and corresponding analysis code are available from the corresponding author upon reasonable request and with appropriate institutional approval.

Funding

This research was funded by the National Science and Technology Council, Taiwan (grant number NSTC 113-2222-E-038-001-MY3) and the Taipei Medical University (grant number TMU111-AE1-B30).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Ye Q, Zou B, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean NAFLD: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2020;5(8):739–752. doi:10.1016/S2468-1253(20)30077-7
2. Cheng P-N, Chen W-J, Hou CJ-Y. Taiwan Association for the Study of the Liver; Taiwan Society of Cardiology. Taiwan position statement on the management of metabolic dysfunction-associated fatty liver disease and cardiovascular diseases. *Clin Mol Hepatol*. 2024;30(1):16–36. doi:10.3350/cmh.2023.0315
3. Li J, Zou B, Yeo YH, et al. Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999–2019: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2019;4(5):389–398. doi:10.1016/S2468-1253(19)30039-1
4. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *J Hepatol*. 2023;79(6):1542–1556. doi:10.1016/j.jhep.2023.06.018
5. Hernaez R, Lazo M, Bonekamp S, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. *Hepatology*. 2011;54(3):1082–1090. doi:10.1002/hep.24452
6. Middleton MS, Van Natta ML, Heba ER, et al. Agreement between MRI-PDFF and histologic hepatic steatosis grade. *Gastroenterology*. 2017;153(3):753–761. doi:10.1053/j.gastro.2017.06.005
7. Stine JG, Munaganuru N, Neuschwander-Tetri BA, Harrison SA, Younossi ZM, Loomba R. Change in MRI-PDFF and histologic response in NASH: systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2021;19(12):2274–2283.e5. doi:10.1016/j.cgh.2020.08.061
8. Park H, Kim SY, Lee SS, et al. Cutoff values of hepatic steatosis for MRI-PDFF using histologic ≥5% fat as reference. *Korean J Radiol*. 2022;23(8):1260–1268. doi:10.3348/kjr.2022.0334
9. Bedogni G, Bellentani S, Miglioli L, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol*. 2006;6:33. doi:10.1186/1471-230X-6-33
10. Lee JH, Kim D, Kim HJ, et al. Hepatic Steatosis Index: a simple screening tool reflecting NAFLD. *Dig Liver Dis*. 2010;42(7):503–508. doi:10.1016/j.dld.2009.08.002
11. Yang BL, Wu WC, Fang KC, et al. External validation of FLI for identifying ultrasonographic fatty liver in a large Taiwanese cohort. *PLoS One*. 2015;10(3):e0120443. doi:10.1371/journal.pone.0120443
12. Zamanian H, Shalbaf A, Zali MR, et al. Application of artificial intelligence techniques for non-alcoholic fatty liver disease diagnosis: a systematic review (2005–2023). *Comput Methods Programs Biomed*. 2024;244:107932. doi:10.1016/j.cmpb.2023.107932
13. Hirooka M, Miyake T, Yano R, et al. Development of a neural network to detect hepatic steatosis in metabolic dysfunction–associated steatotic liver disease. *Gastro Hep Adv*. 2025;5(1):100765. doi:10.1016/j.gastha.2025.100765
14. Yin C, Zhang H, Du J, Zhu Y, Zhu H, Yue H. Artificial intelligence in imaging for liver disease diagnosis. *Front Med*. 2025;12:1591523. doi:10.3389/fmed.2025.1591523
15. Qin S, Hou X, Wen Y, et al. Machine learning classifiers for screening nonalcoholic fatty liver disease in general adults. *Sci Rep*. 2023;13:3638. doi:10.1038/s41598-023-30750-5
16. Peng HY, Duan SJ, Pan L, et al. Development and validation of machine learning models for nonalcoholic fatty liver disease. *Hepatobiliary Pancreat Dis Int*. 2023;22:615–621. doi:10.1016/j.hbpd.2023.03.009
17. Zhang H, Liu J, Su D, et al. Diagnostic of fatty liver using radiomics and deep learning models on non-contrast abdominal CT. *PLoS One*. 2025;20(2):e0310938. doi:10.1371/journal.pone.0310938
18. Schattenberg JM, Balp -M-M, Reinhart B, et al. NASHmap: clinical utility of a machine learning model to identify patients at risk of NASH in real-world settings. *Sci Rep*. 2023;13(1):5573. doi:10.1038/s41598-023-32551-2
19. Byra M, Styczynski G, Szmigielski C, et al. Transfer learning with CNNs for liver steatosis assessment in ultrasound images. *Int J Comput Assist Radiol Surg*. 2018;13(12):1895–1903. doi:10.1007/s11548-018-1843-2
20. Kim T, Lee J, Kwon H, et al. Deep learning on multi-view ultrasound for fatty liver detection and fat-fraction estimation across scanners. *JMIR Med Inform*. 2021;9(11):e30066. doi:10.2196/30066
21. Wang K, Yokoo T, Cui J, et al. Deep learning inference of hepatic PDFF from routine T1-weighted MR (in-/opposed-phase) using CSE-MRI PDFF as reference. *AJR Am J Roentgenol*. 2023. doi:10.2214/AJR.23.29607
22. Ma H, Xu C-F, Shen Z, Yu C-H, Li Y-M. Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. *Biomed Res Int*. 2018;2018:4304376. doi:10.1155/2018/4304376
23. Yip TCF, Ma AJ, Wong VWS, et al. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease in the general population. *Aliment Pharmacol Ther*. 2017:1–10. doi:10.1111/apt.14172.

24. Atabaki-Pasdar N, Ohlsson M, Viñuela A, et al. Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the IMI DIRECT cohorts. *PLoS Med*. 2020;17(6):e1003149. doi:10.1371/journal.pmed.1003149
25. Wu -C-C, Yeh W-C, Hsu W-D, et al. Prediction of fatty liver disease using machine learning algorithms. *Comput Methods Programs Biomed*. 2019;170:23–29. doi:10.1016/j.cmpb.2018.12.032
26. Li J, Chen J, Zeng X, Lyu G, Lin S, He S. Update of machine learning for ultrasound diagnosis of metabolic dysfunction-associated steatotic liver disease: a bright future for deep learning. *PeerJ*. 2025;13:e19645. doi:10.7717/peerj.19645
27. Zhang Y, Liu X, Zhang X, Fei Y, Li X. Machine learning-based prediction of metabolic dysfunction-associated steatotic liver disease using National Health and Nutrition Examination Survey (NHANES) data. *PLoS One*. 2025;20(11):e0335656. doi:10.1371/journal.pone.0335656
28. Nabrdalik K, Kwiendacz H, Irlik K, et al. Machine learning identifies metabolic dysfunction–associated steatotic liver disease in patients with diabetes mellitus. *J Clin Endocrinol Metab*. 2024;109(8):2029–2038. doi:10.1210/clinem/dgae060
29. McTeer M, Applegate D, Mesenbrink P, et al. Machine learning approaches to enhance diagnosis and staging of patients with MASLD using routinely available clinical information. *PLoS One*. 2024;19(2):e0299487. doi:10.1371/journal.pone.0299487
30. Song K, Kwon YJ, Lee E, et al. Machine learning-based model for predicting metabolic dysfunction-associated steatotic liver disease using non-invasive parameters in young adults. *Front Endocrinol*. 2025;16:1701729. doi:10.3389/fendo.2025.1701729
31. Sterling RK, Lissen E, Clumeck N, et al. FIB-4 index. *Hepatology*. 2006;43(6):1317–1325. doi:10.1002/hep.21178
32. Angulo P, Hui JM, Marchesini G, et al. NAFLD Fibrosis Score. *Hepatology*. 2007;45(4):846–854. doi:10.1002/hep.21496
33. Lundberg SM, Lee SI. A unified approach to interpreting model predictions (SHAP). *Adv Neural Inf Process Syst*. 2017;arXiv:1705.07874.