


RESEARCH

Open Access



Machine learning models for predicting metabolic dysfunction-associated steatotic liver disease prevalence using basic demographic and clinical characteristics

Gangfeng Zhu¹, Yipeng Song¹, Zenghong Lu², Qiang Yi¹, Rui Xu³, Yi Xie¹, Shi Geng⁴, Na Yang^{4,5}, Liangjian Zheng¹, Xiaofei Feng², Rui Zhu², Xiangcai Wang^{2*}, Li Huang^{2*} and Yi Xiang^{2,6*} 

Abstract

Background Metabolic dysfunction-associated steatotic liver disease (MASLD) is a global health concern that necessitates early screening and timely intervention to improve prognosis. The current diagnostic protocols for MASLD involve complex procedures in specialised medical centres. This study aimed to explore the feasibility of utilising machine learning models to accurately screen for MASLD in large populations based on a combination of essential demographic and clinical characteristics.

Methods A total of 10,007 outpatients who underwent transient elastography at the First Affiliated Hospital of Gannan Medical University were enrolled to form a derivation cohort. Using eight demographic and clinical characteristics (age, educational level, height, weight, waist and hip circumference, and history of hypertension and diabetes), we built predictive models for MASLD (classified as none or mild: controlled attenuation parameter (CAP) ≤ 269 dB/m; moderate: 269–296 dB/m; severe: CAP > 296 dB/m) employing 10 machine learning algorithms: logistic regression (LR), multilayer perceptron (MLP), extreme gradient boosting (XGBoost), bootstrap aggregating, decision tree, K-nearest neighbours, light gradient boosting machine, naive Bayes, random forest, and support vector machine. These models were externally validated using the National Health and Nutrition Examination Survey (NHANES) 2017–2023 datasets.

Results In the hospital outpatient cohort, machine learning algorithms demonstrated robust predictive capabilities. Notably, LR achieved the highest accuracy (ACC) of 0.711 in the test cohort and 0.728 in the validation cohort, coupled with robust areas under the receiver operating characteristic curve (AUC) values of 0.798 and 0.806, respectively. Similarly, MLP and XGBoost showed promising results, with MLP achieving an ACC of 0.735 in the test

*Correspondence:
Xiangcai Wang
wangxiangcai@cscso.ac.cn
Li Huang
hlhellen@gmu.edu.cn
Yi Xiang
xiangyi_xiangyi@126.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

cohort, and XGBoost registering an AUC of 0.798. External validation using the NHANES datasets yielded consistent AUC results, with LR (0.831), MLP (0.823), and XGBoost (0.784) performing robustly.

Conclusions This study demonstrated that machine learning models constructed using a combination of essential demographic and clinical characteristics can accurately screen for MASLD in the general population. This approach significantly enhances the feasibility, accessibility, and compliance of MASLD screening and provides an effective tool for large-scale health assessments and early intervention strategies.

Keywords Metabolic dysfunction-associated steatotic liver disease, Demographic and clinical characteristics, Machine learning, Non-invasive screening, National health and nutrition examination survey

Background

Metabolic dysfunction-associated steatotic liver disease (MASLD), characterised by the accumulation of excess fat in the liver of individuals who consume little or no alcohol, has emerged as a global public health concern [1]. Its epidemiology reflects a troubling upward trend, mirroring the increasing prevalence of obesity and diabetes mellitus (DM) [2, 3]. As the most common chronic liver condition in the Western world, MASLD encompasses a spectrum of liver disorders, ranging from simple steatosis to non-alcoholic steatohepatitis, potentially progressing to cirrhosis and hepatocellular carcinoma [4, 5, 6]. The insidious nature of MASLD, of being often asymptomatic in its early stages, underscores the critical need for its early detection and intervention [7, 8, 9]. Furthermore, the multifactorial aetiology of the disease, involving complex interactions between genetic predisposition, lifestyle factors, and metabolic syndrome, poses significant challenges to its management and risk stratification [10, 11]. Thus, early and accurate risk assessment of MASLD is not only pivotal for timely therapeutic interventions but also paramount in preventing its progression to more advanced liver diseases [12, 13].

Contemporary diagnostic modalities for MASLD predominantly pivot around transient elastography (TE), a technique acclaimed for its diagnostic precision. TE, leveraging its ability to quantify liver stiffness, serves as a pivotal instrument in not only diagnosing MASLD but also in stratifying the disease severity [14, 15]. Despite its efficacy, TE relies on sophisticated equipment and the necessity of a clinical setting for implementation poses constraints, especially in terms of accessibility and patient compliance [16]. Similarly, traditional diagnostic approaches, including clinical evaluations, biochemical markers, and liver biopsies, underscore the dependency on patient cooperation and the need for specialised medical infrastructure [17]. Such methodologies, although integral, often fall short of convenience and patient adherence, impeding widespread screening and early detection efforts [18]. The inherent limitations of these approaches underscore the need for alternative and more accessible diagnostic strategies that can circumvent the

forementioned barriers and foster enhanced patient engagement in disease management [19].

The burgeoning field of machine learning (ML) presents a paradigm shift in medical diagnostics, offering a potent amalgam of precision and adaptability that traditional methodologies often lack [20, 21]. ML algorithms exhibit substantial potential in the realm of MASLD, primarily for refining risk stratification and prognostic assessment [22]. These algorithms can adeptly analyse complex multidimensional datasets and transcend the limitations of human cognition. In particular, methods such as random forest (RF), support vector machines (SVM), and neural networks have demonstrated remarkable capabilities in discerning intricate patterns and interactions among clinical variables [23]. This process not only enhances predictive accuracy but also facilitates the identification of novel risk factors and patient subgroups, thus enabling a more nuanced understanding of MASLD pathophysiology [24]. The adoption of ML in MASLD risk assessment could revolutionise the approach for early detection, allowing for timely intervention and personalised management strategies [25]. Moreover, the ability of ML models to process and interpret large-scale data swiftly aligns seamlessly with the demands of contemporary clinical practice, heralding a new era in the efficient and effective management of MASLD [24, 26].

Several recent studies have demonstrated promising applications of ML models in the screening and diagnosis of MASLD, yielding encouraging results. For instance, Chen et al. developed a risk prediction model based on multiple potential indicators, including homeostasis model assessment of insulin resistance (HOMA-IR), triglyceride glucose-waist circumference (TyG-WC), and achieved an impressive area under the receiver operating characteristic curve (AUC) of 0.960 [27]. Similarly, McTeer et al. used European MASLD registry data to construct various ML models using conventional clinical parameters, with AUC values ranging from 0.719 to 0.994, highlighting the utility of readily available clinical information in predicting MASLD stages and outcomes [28]. Nabrdalik et al. focused on ML-based recognition methods for MASLD in patients with diabetes, constructed a model using eight core parameters, and

achieved an AUC of 0.84, demonstrating the potential of the model for cardiovascular risk prevention [29]. Despite the advancements presented by these studies, reliance on complex clinical or laboratory indicators, such as HOMA-IR and lipid-related parameters, limits the feasibility of MASLD screening, particularly in resource-limited settings or for large-scale population assessments. Although these advanced biomarkers are valuable in clinical practice, they are often not readily accessible for general screening. In contrast, this study adopted an innovative approach using a minimal set of easily accessible demographic and clinical factors, such as age, height, and weight, to develop a ML-based screening model. This strategy enhances the scalability and accessibility of MASLD detection, making it more applicable in diverse healthcare settings, improving early detection in underserved populations and facilitating timely stratified diagnosis, treatment, and intervention.

Leveraging ML models built on easily obtainable parameters, such as age, sex, body mass index (BMI), and chronic disease history, has emerged as a pivotal strategy to enhance accessibility and patient compliance in MASLD risk assessment [30]. This approach democratises the diagnostic process by transcending the limitations of clinical settings and specialised equipment. By integrating these readily available and self-measurable indicators within an ML framework, the potential to substantially increase diagnostic convenience and patient adherence can be realised. Furthermore, the adoption of such models in community settings amplifies screening ubiquity, making MASLD risk assessment more inclusive and accessible to the general populace [31, 32]. This paradigm fosters early detection through widespread accessibility and aligns with patient-centred care principles. The operational simplicity and non-reliance on invasive procedures inherent in this methodology substantially reduce barriers to screening, thereby paving the way for timely interventions and mitigating the progression of MASLD. Embracing this innovative approach could mark a significant milestone in MASLD management, aligning with the contemporary shifts towards preventative health strategies and personalised medicine.

This study investigated the efficacy of 10 ML algorithms in predicting MASLD prevalence using a combination of basic demographic characteristics and essential clinical indicators, including age, educational level, height, weight, waist and hip circumference, and history of hypertension and diabetes. Using a comprehensive outpatient cohort and external validation with National Health and Nutrition Examination Survey (NHANES) data, we sought to establish a feasible and accessible approach for MASLD screening in large populations.

Methods

Study design and participants

This prospective study, conducted at the First Affiliated Hospital of Gannan Medical University, enrolled 10,007 outpatients between October 2020 and February, 2024. Eligible participants were aged 15 years or older and of any sex. The participants provided informed consent after understanding the study's objectives, methodology, and importance. Individuals were excluded if they had severe cardiovascular or cerebrovascular diseases or malignant tumours, were pregnant, or were unable to adhere to the study protocol, including completing questionnaires and undergoing TE (Fig. 1). The study was approved by the relevant medical ethics committee and all participants provided written informed consent.

Demographic data collection

Comprehensive data collection involved direct interviews and questionnaires, capturing demographic details (age, sex, educational level, and occupation); a thorough review of diabetes diagnoses, including type, duration, and treatment modalities; and medical history with a focus on hepatitis B virus (HBV) and hepatitis C virus (HCV) infections and other liver diseases. Physical assessments included height, weight, BMI and hypertension status. Detailed definitions and descriptions of all study variables are provided in Table S1.

MASLD assessment

Liver fat alterations were evaluated using FibroScan (Echosens, France) for TE. This non-invasive technique measures the controlled attenuation parameter (CAP) to quantitatively evaluate liver fat content [33]. Participants fasted for at least 3 h prior to the examination. Trained and certified medical technicians operated the FibroScan TE device to measure the CAP [34]. Measurements were performed with the participants in the supine position, with their right arm behind the head, and a slight leftward tilt of the upper body to maximally expand the intercostal space. The measurement area ranged from the right anterior axillary line to the mid-axillary line between the seventh and eighth, or eighth and ninth ribs, maintaining a probe orientation perpendicular to the skin surface. Quality control standards for each participant required at least 10 valid CAP measurements, with a quartile-to-median ratio $\leq 30\%$, taking the median value as the final measurement.

CAP was used to evaluate liver cell fat content. MASLD was classified as follows: none or mild: $\text{CAP} \leq 269 \text{ dB/m}$; moderate: $269\text{--}296 \text{ dB/m}$; severe: $\text{CAP} > 296 \text{ dB/m}$.

Data preprocessing and feature selection

In this study, the missForest algorithm was employed to handle the missing data ($<25\%$) in both the original and

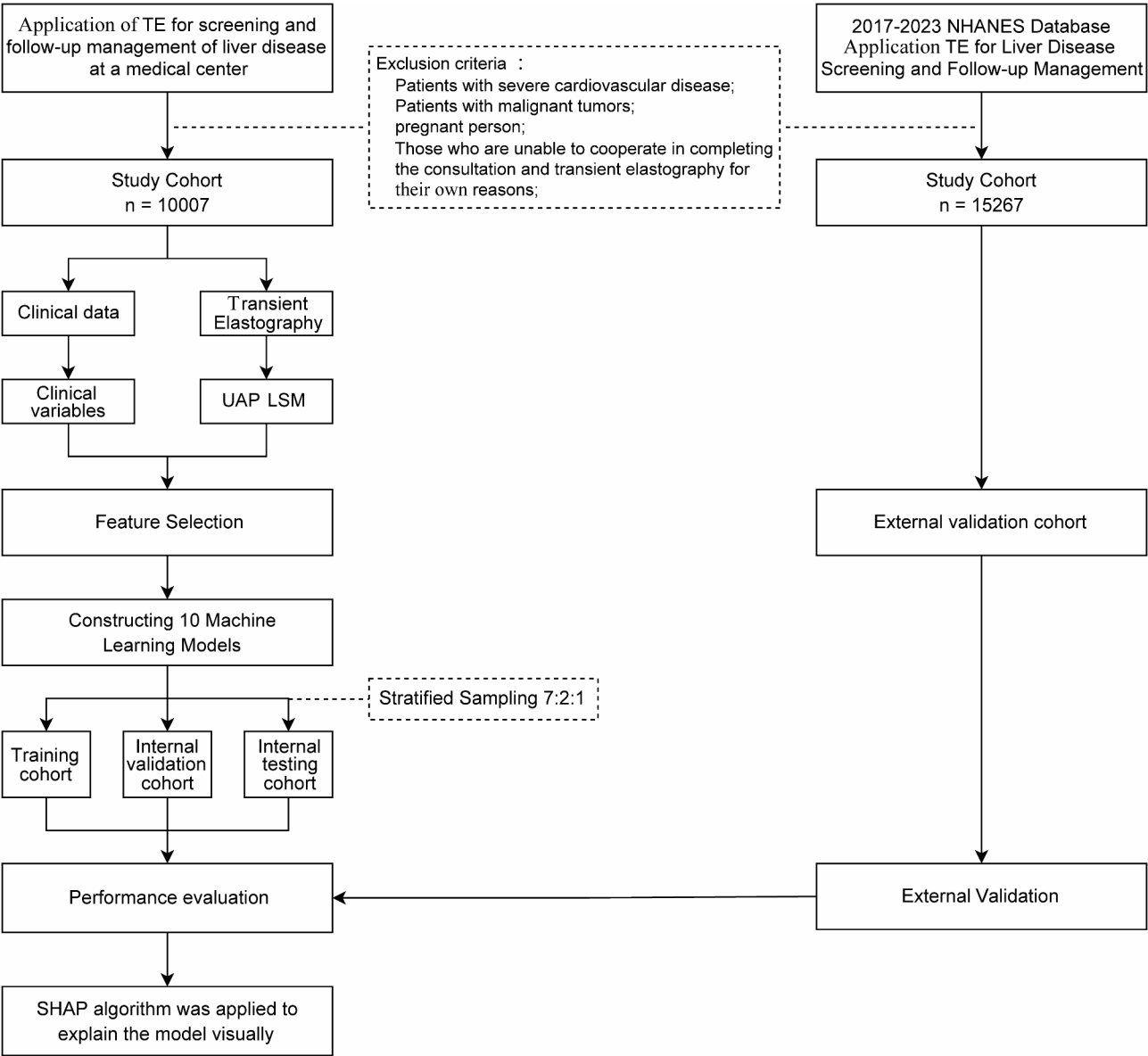


Fig. 1 Flowchart of this study

NHANES datasets. This approach efficiently handles both continuous and categorical variables by leveraging the observed values from other variables to predict the missing variables. By imputing missing data, missForest preserves data completeness and ensures the reliability of subsequent analyses. Table S2 summarises the proportion of missing data for each variable, and the corresponding imputation methods applied to both datasets.

For feature selection, we employed a two-step approach combining least absolute shrinkage and selection operator (LASSO) regression and the Boruta algorithm. LASSO regression, which is effective in handling multicollinearity and enhancing model interpretability, was applied to refine the variables by reducing the coefficients

of less informative predictors to zero. The selection process was guided by the mean squared error and alpha values to optimise the model performance (Figure S1A–B). Subsequently, the Boruta algorithm, a robust and stable method, was used to identify relevant variables by comparing the importance of predictive variables with that of randomly permuted ‘shadow variables’ through multiple iterations of the RF algorithm, systematically excluding non-essential variables (Figure S1C). The intersection of the variables selected using both methods was used as the foundation for subsequent analyses.

ML algorithms and model development

To predict MASLD risk stratification, we employed 10 state-of-the-art ML tools for developing predictive models: logistic regression (LR), multilayer perceptron (MLP), extreme gradient boosting (XGBoost), bootstrap aggregating (Bagging), decision tree (DT), K-nearest neighbours (KNN), light gradient boosting machine (LightGBM), naive Bayes (NB), RF, and SVM. The dataset was divided into training (70%), validation (20%), and test (10%) sets. All analyses were conducted using Python 3.8.10 and R 3.3.2. LR, a multilinear regression method, utilises maximum likelihood estimation to gauge predictive probabilities. MLP, a neural network-based algorithm, captures complex non-linear relationships through interconnected layers and optimises weights via backpropagation. Bagging improves stability and accuracy by generating multiple training subsets through resampling and aggregating predictions from individual models. KNN classifies instances by measuring the proximity in the feature space, relying on majority voting among the nearest neighbours to make predictions. NB, which is a probabilistic model based on Bayes' theorem, assumes feature independence and demonstrates robust performance with categorical and continuous data. SVM classifies data points by maximising the margin in a transformed dimensional space. DT is a structured tree predictive model that selects optimal features and splits points based on the Gini index. RF, which employs DTs as base classifiers, recursively builds multiple trees, with the final prediction determined by a voting mechanism. XGBoost, another tree-based ensemble classifier, minimises the current loss function by generating new decision trees until the residual of the loss function approaches zero, by employing techniques such as second-order derivatives, Taylor series, and parallel learning. LightGBM, an advanced gradient boosting DT algorithm, enhances the performance through parameter optimisation, including adjustments in iteration numbers, maximum leaves, and tree depth. After the initial development of the MASLD risk prediction models, hyperparameter tuning was conducted for all 10 ML algorithms to optimise model performance. Specifically, Bayesian optimisation was selectively applied to models with a higher number of hyperparameters and more complex tuning requirements, including XGBoost, LightGBM, RF, SVM, and MLP, owing to its efficiency in navigating high-dimensional parameter spaces. For simpler models, such as LR, NB, DT, KNN, and Bagging, hyperparameters were optimised using grid search or random search because their parameter spaces were relatively limited and computationally less intensive. The final hyperparameters adopted for each model are listed in Table S3.

Prediction model evaluation

To comprehensively assess the performance of the ML models in predicting MASLD risk, a suite of evaluation metrics was employed. The primary metric used was the AUC, which provides a measure of the model's ability to distinguish between different risk categories. A higher AUC value indicates a greater discriminatory power of a model. Additional metrics included accuracy (ACC, the proportion of true results among the total number of cases examined), sensitivity (ability of the model to correctly identify individuals with MASLD), specificity (ability of the model to correctly identify individuals without MASLD), precision (the proportion of true positive predictions among all positive predictions, reflecting the reliability of the model in identifying individuals with MASLD), and F1 score (a harmonic mean of precision and recall, providing a balance between sensitivity and specificity). These metrics collectively offer a multifaceted view of the effectiveness of the models, ensuring a thorough and nuanced evaluation of their predictive capabilities in the context of MASLD risk stratification.

External validation using NHANES data

NHANES data from a comprehensive cross-sectional population-based study were utilised to supplement our findings. By covering the cycles from 2017 to 2023, the NHANES provides crucial insights into the health and nutritional status of the U.S. population. The NHANES protocol, approved by the National Center for Health Statistics Research Ethics Review Board, ensured rigorous ethical standards, and all participants provided written informed consent. The NHANES data from 2017 to 2023 were used for external validation. We extracted demographic data (age, sex, and educational level) and disease history (hypertension and DM) from the NHANES data. The physical examination metrics included height, weight, BMI, waist and hip circumference, liver stiffness measurement, and CAP. Detailed mapping and descriptions of the variables used in this study are provided in Table S4. This table outlines the variable codes, predictors, descriptions, types, and value ranges, ensuring traceability and comparability of the data across the NHANES datasets.

Statistical analysis

All statistical analyses were conducted using SPSS software (version 23.0), R software (version 3.3.2), and Python software (version 3.8.10). Continuous variables were presented as mean \pm standard deviation for normally distributed data and as median with interquartile range for non-normally distributed data. Categorical variables were summarised as percentages and frequencies. Comparative analyses between the groups were performed using non-parametric tests for continuous

variables (Mann–Whitney U test), as all continuous data in this study were non-normally distributed. For categorical variables, the chi-squared test was used to compare group differences. Statistical significance was set at $P < 0.05$.

Results

Demographic characteristics of participants

This study enrolled 10,007 outpatients who were categorised into distinct groups: a training cohort comprising 7,003 individuals, an internal validation cohort with 2,002 participants, and an internal testing cohort containing 1,002 participants. The allocation followed a stratified randomisation approach, at a ratio of 7:2:1, ensuring that each subgroup retained a similar distribution of relevant features. Detailed descriptions of the baseline demographics, physical evaluations, and clinical diagnoses of each cohort are summarised in Table 1. In the evaluation of the demographic variables, significant disparities emerged between the MASLD and non-MASLD groups. For ease of reference, detailed data are presented in Table S5. The median age was higher in the MASLD group (52 [40–58] years) than in the non-MASLD group (49 [37–57] years; $P < 0.001$). Sex distribution also differed significantly, with female individuals comprising 56.21% ($n = 3235$) and 68.72% ($n = 2922$) of the MASLD and non-MASLD groups, respectively ($P < 0.001$). Education level showed a notable variation between the groups, with a higher proportion of individuals with low education (12.27% vs. 9.05%) and a lower proportion of those with higher education (25.49% vs. 29.70%, $P < 0.001$) in the MASLD group compared with the non-MASLD group. Occupational distribution further highlighted the intergroup differences, with a slightly higher percentage of the labour force in the MASLD group than in the non-MASLD group (77.15% vs. 75.16%, $P = 0.008$). Physical examination data demonstrated that the MASLD group exhibited significantly higher median values for height (164 [158–170] cm), weight (67 [60–75] kg), BMI (24.91 [23.23–27.06] kg/m²), waist circumference (85 [80–91] cm), and hip circumference (97 [90–101] cm), compared with the non-MASLD group (all $P < 0.001$). Additionally, the MASLD group exhibited a higher prevalence of DM (4.07%) than the non-MASLD group did (3.10%, $P = 0.013$), further highlighting the association between MASLD and DM. Similarly, hypertension was noted significantly more frequently in the MASLD group (27.72%) than in the non-MASLD group (17.03%, $P < 0.001$), highlighting the metabolic and cardiovascular risk profile associated with MASLD. Notably, the prevalence of HBV and HCV infections did not differ significantly between the two groups; HBV was observed in 1.60% of the MASLD cohort and 1.70% of the non-MASLD cohort ($P = 0.828$) and HCV in 0.10% and 0.20%, respectively

($P = 0.528$). These findings collectively underscore the distinctive demographic and clinical profiles of MASLD and non-MASLD populations.

Feature selection

The two-step feature selection approach identified a robust predictor set for the MASLD. LASSO regression and the Boruta algorithm yielded a combined set of statistically significant variables, including age, educational level, height, weight, BMI, waist and hip circumference, and history of hypertension and DM (Figure S1D). Notably, although BMI was initially identified as relevant, it was excluded from the final model because of its high correlation with height and weight, which could lead to multicollinearity issues in ML modelling.

Development of ML models for MASLD prediction

We employed 10 ML algorithms based on demographic characteristics to construct diagnostic models for MASLD. These include LR, MLP, XGBoost, Bagging, RF, NB, LightGBM, SVM, KNN, and DT. To further investigate the contribution of each feature across the different models, we assessed the variable importance rankings derived from all 10 ML algorithms (Figure S2). LR emerged as the most effective, achieving an ACC of 0.728, a sensitivity of 0.749, and an AUC of 0.806 in the internal validation cohort (Fig. 2A; Table 2). XGBoost and MLP also exhibited commendable performances, with XGBoost slightly trailing LR in terms of sensitivity. The Bagging model, although comparable in ACC to the LR model, had a lower sensitivity. Conversely, DT demonstrated the least effectiveness, with the lowest ACC and sensitivity, underscoring its limitations. In the internal testing cohort, LR maintained its lead, with an ACC of 0.711 and an AUC of 0.798. Other models, such as XGBoost and MLP, also demonstrated robust performance, with AUCs of 0.798 and 0.796, respectively (Fig. 2B). To further assess the classification efficacy of the LR model, confusion matrices were generated for both the validation and test cohorts (Figures S3A–B). These matrices provide a detailed breakdown of true positives, false positives, true negatives, and false negatives for the model. The findings indicate that the LR model effectively identified individuals with MASLD (true positives) and those without MASLD (true negatives). Notably, the relatively low incidence of false positives and false negatives underscore the reliability of the model's clinical risk prediction, minimising the likelihood of both overdiagnosis and missed cases. These results align with the model's strong performance metrics, including high AUC and F1 scores, further reinforcing its potential for clinical application. Additionally, pairwise comparisons of the AUCs between LR and alternative classification models, as detailed in Tables S6–S8, substantiate the statistical

Table 1 Baseline characteristics and outcomes of MASLD patients in the training, internal validation and internal testing cohort

Characteristics	Training cohort (n = 7003)	Internal validation cohort (n = 2002)	P-value	Internal testing cohort (n = 1002)
Demographic characteristics				
Age (years)	51 (39–58)	51 (39–58)	0.279	52 (40–58)
Sex, n (%)			1.000	
Female	4299 (61.39)	1229 (61.39)		629 (62.77)
Male	2704 (38.61)	773 (38.61)		373 (37.23)
Educational, n (%)			0.483	
Low education	752 (10.74)	222 (11.09)		117 (11.68)
Secondary education	4292 (61.29)	1247 (62.29)		647 (64.57)
Higher education	1959 (27.97)	533 (26.62)		238 (23.75)
Occupation, n (%)			0.808	
Labor force	5320 (75.97)	1524 (76.12)		792 (79.04)
Corporate/Government sector	1659 (23.69)	473 (23.63)		207 (20.66)
Student group	24 (0.34)	5 (0.25)		3 (0.30)
Height (cm)	163 (158–170)	163 (158–170)	0.527	162.75 (158.00–170.00)
Weight (cm)	64 (57–70)	63 (57–70)	0.231	63 (56–70)
BMI (kg/m ²)	23.78 (21.80–25.81)	23.67 (21.63–25.80)	0.376	23.67 (21.72–25.78)
Waist circumference (cm)	83 (80–90)	82 (80–89)	0.065	83 (79–90)
Hip circumference (cm)	95 (90–100)	95 (90–100)	0.580	95 (90–100)
Medical history				
Hypertension, n (%)			0.977	
No history of hypertension	5386 (76.91)	1541 (76.97)		761 (75.95)
History of hypertension	1617 (23.09)	461 (23.03)		241 (24.05)
DM, n (%)			0.508	
No history of DM	6751 (96.40)	1923 (96.05)		967 (96.51)
History of DM	252 (3.60)	79 (3.95)		35 (3.49)
HBV, n (%)			0.107	
No history of HBV	6901 (98.54)	1962 (98.00)		981 (97.90)
History of HBV	102 (1.46)	40 (2.00)		21 (2.10)
HCV, n (%)			1.000	
No history of HCV	6993 (99.86)	2000 (99.90)		1000 (99.80)
History of HCV	10 (0.14)	2 (0.10)		2 (0.20)
Liver Disease Characteristics				
LSM (kPa)	6.10 (5.40–7.10)	6.10 (5.30–7.00)	0.347	6.1 (5.30–7.10)
cACLD, n (%)			0.970	
No history of cACLD	6667 (95.20)	1907 (95.25)		946 (94.41)
History of cACLD	336 (4.80)	95 (4.75)		56 (5.59)
Liver Fibrosis, n (%)			0.533	
No history of liver fibrosis	5419 (77.38)	1563 (78.07)		775 (77.35)
History of liver fibrosis	1584 (22.62)	439 (21.93)		227 (22.65)
CAP (dB/m)	254 (226–274)	253 (228–274)	0.955	252.50 (227.00–274.00)
Outcome measurements				
MASLD, n (%)			1.000	
No history of MASLD	2975 (42.48)	851 (42.51)		426 (42.51)
History of MASLD	4028 (57.52)	1151 (57.49)		576 (57.49)

Note: Continuous variables are presented as median (IQR). Categorical variables are presented as n (%).

Abbreviations: BMI: body mass index; DM: diabetes mellitus; HBV: hepatitis B virus; HCV: hepatitis C virus; LSM: liver stiffness measurement; cACLD: compensated alcoholic chronic liver disease; CAP: controlled attenuation parameter; MASLD: metabolic dysfunction-associated steatotic liver disease

significance of performance differences, underscoring the robustness of the LR model in predictive accuracy.

A comprehensive decision curve analysis (DCA) conducted on the internal validation cohort demonstrated the superior clinical utility of the LR model among

various ML approaches, reinforcing its applicability in real-world settings (Figure S4A). Similarly, DCA results for the internal testing cohort confirmed that using the LR model for risk prediction yielded substantial net benefits (Figures S4B). Figure S5A and B depict the calibration

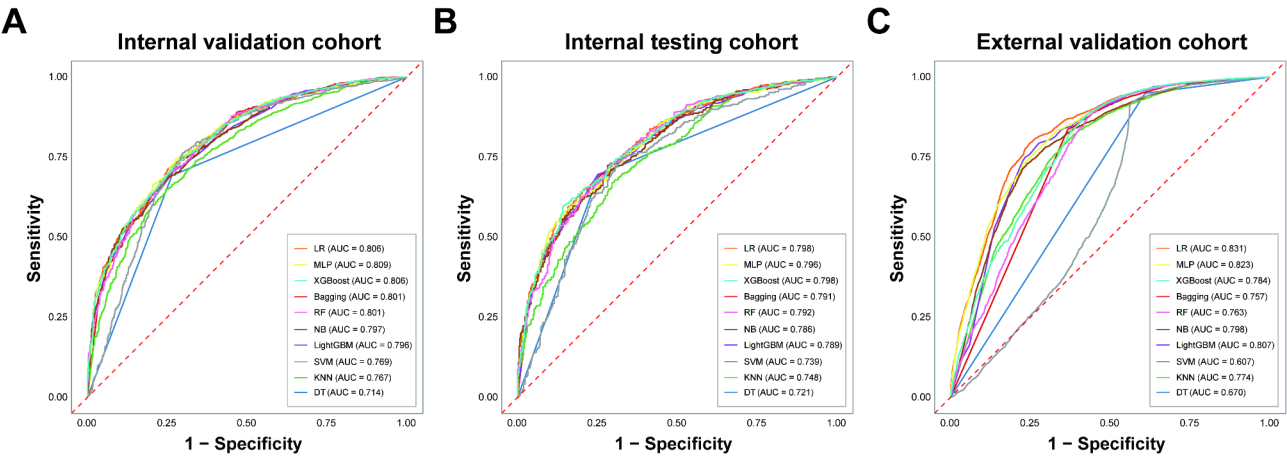


Fig. 2 Prediction performance by AUCs for the ten machine-learning models in the internal validation cohort (A), internal testing cohort (B) and external validation cohort (C)

Table 2 Performance of various machine learning models for predicting MASLD across internal validation, internal testing, and external validation cohorts

Cohort	Model	AUC	95% CI	Optimal cutoff	Accuracy	Sensitivity	Specificity	Precision	F1 score
Internal validation cohort	LR	0.806	(0.787–0.825)	0.571	0.728	0.749	0.713	0.659	0.701
	MLP	0.809	(0.790–0.827)	0.531	0.735	0.724	0.744	0.676	0.699
	XGBoost	0.806	(0.788–0.825)	0.566	0.727	0.730	0.725	0.662	0.694
	Bagging	0.801	(0.782–0.820)	0.578	0.727	0.709	0.740	0.669	0.688
	RF	0.801	(0.782–0.820)	0.578	0.719	0.750	0.697	0.646	0.694
	NB	0.797	(0.778–0.816)	0.464	0.714	0.747	0.689	0.640	0.689
	LightGBM	0.796	(0.776–0.815)	0.577	0.716	0.731	0.705	0.647	0.687
	SVM	0.769	(0.747–0.791)	0.594	0.736	0.704	0.760	0.685	0.694
	KNN	0.767	(0.746–0.788)	0.619	0.694	0.753	0.650	0.614	0.677
	DT	0.714	(0.694–0.734)	0.569	0.711	0.736	0.692	0.639	0.684
Internal testing cohort	LR	0.798	(0.771–0.825)	0.617	0.711	0.793	0.649	0.626	0.700
	MLP	0.796	(0.769–0.823)	0.579	0.697	0.847	0.585	0.602	0.704
	XGBoost	0.798	(0.771–0.825)	0.641	0.709	0.857	0.599	0.612	0.714
	Bagging	0.791	(0.763–0.818)	0.597	0.722	0.718	0.724	0.658	0.687
	RF	0.792	(0.764–0.819)	0.556	0.720	0.700	0.734	0.661	0.680
	NB	0.786	(0.759–0.814)	0.664	0.695	0.826	0.597	0.603	0.697
	LightGBM	0.789	(0.762–0.817)	0.582	0.719	0.749	0.696	0.646	0.693
	SVM	0.739	(0.707–0.770)	0.610	0.717	0.674	0.748	0.664	0.669
	KNN	0.748	(0.718–0.778)	0.614	0.685	0.704	0.670	0.612	0.655
	DT	0.721	(0.693–0.749)	0.569	0.718	0.746	0.696	0.645	0.692
External validation cohort	LR	0.831	(0.825–0.837)	0.923	0.777	0.756	0.790	0.695	0.724
	MLP	0.823	(0.817–0.829)	0.691	0.755	0.734	0.769	0.667	0.699
	XGBoost	0.784	(0.777–0.791)	0.788	0.763	0.574	0.883	0.755	0.652
	Bagging	0.757	(0.751–0.764)	0.874	0.758	0.633	0.837	0.710	0.669
	RF	0.763	(0.755–0.770)	0.726	0.756	0.571	0.873	0.740	0.645
	NB	0.798	(0.792–0.805)	0.997	0.742	0.762	0.729	0.640	0.696
	LightGBM	0.807	(0.800–0.814)	0.912	0.766	0.723	0.794	0.689	0.706
	SVM	0.607	(0.598–0.615)	0.641	0.731	0.436	0.917	0.769	0.556
	KNN	0.774	(0.767–0.781)	0.880	0.738	0.628	0.807	0.673	0.650
	DT	0.670	(0.664–0.676)	0.569	0.733	0.391	0.949	0.829	0.532

Abbreviations: LR, logistic regression; MLP, multilayer perceptron; XGBoost, extreme gradient boosting; Bagging, bootstrap aggregating; DT, decision tree; KNN, k-nearest neighbors; LightGBM, light gradient boosting machine; NB, naive bayes; RF, random forest; SVM, support vector machine; AUC, area under the curve; MASLD, metabolic dysfunction-associated steatotic liver disease

curves for different models in the internal validation and testing cohorts. The LR model exhibited strong calibration for all datasets, with its calibration curves closely aligning with the ideal 45° reference line, indicating high concordance between predicted and actual event probabilities. In contrast, the DT and NB models displayed pronounced calibration biases, particularly in higher probability ranges, where their predictions deviated markedly from observed outcomes.

Furthermore, the analysis of the precision-recall (PR) curves across the models (Figure S6A–B) further highlights the superior performance of the LR model. PR curves are particularly useful for assessing models on imbalanced datasets, as they emphasize the trade-off between precision (positive predictive value) and recall (sensitivity). The LR model demonstrated high precision while maintaining strong recall, effectively minimizing false positives without compromising sensitivity. These findings underscore the robust discriminatory capability and reliability of the LR model for clinical risk prediction in imbalanced clinical datasets, making it the optimal choice for this application.

External validation of the predictive models

We used the NHANES dataset (2017–2023) as an external validation cohort to evaluate the predictive performance of the models. A comprehensive baseline analysis was performed to compare demographic, clinical, and liver-related characteristics between the MASLD and non-MASLD groups (Table S9). Key findings indicated that patients with MASLD exhibited significant differences across multiple dimensions, compared with individuals without MASLD. Demographically, patients with MASLD were older (median age: 55 vs. 40 years, $P < 0.001$), more likely to be male (50.73% vs. 45.16%, $P < 0.001$), and had higher rates of low or secondary education ($P < 0.001$), compared with individuals without MASLD. Anthropometric measurements further highlighted the substantial disparities. Patients with MASLD had significantly elevated BMI (median: 31.10 vs. 24.50 kg/m², $P < 0.001$), waist circumference (105.90 vs. 86.90 cm, $P < 0.001$), and hip circumference (109.10 vs. 98.30 cm, $P < 0.001$), compared with individuals without MASLD. These differences were accompanied by higher weight and height, underscoring the metabolic profile of MASLD. Clinically, patients with MASLD demonstrated a markedly higher prevalence of hypertension (42.20% vs. 22.94%, $P < 0.001$) and DM (17.94% vs. 5.87%, $P < 0.001$).

Post-imputation analysis (Table S10) corroborated these findings, with similar trends observed for demographics and metabolism. Patients with MASLD were significantly older, had a higher BMI and waist circumference, and had a greater prevalence of related conditions, compared with individuals without MASLD. These

results provide strong evidence of the distinct demographic and clinical profiles of patients with MASLD, emphasising the importance of these variables in predictive modelling and their relevance for external validation using real-world datasets.

External validation confirmed the robustness and generalisability of our predictive models, with LR emerging as the best-performing model. The LR model achieved an AUC of 0.831, outperforming the other models, as demonstrated by the receiver operating characteristic curves (Fig. 2C). This indicated a strong ability to distinguish patients with MASLD from individuals without MASLD in the NHANES dataset. We conducted a series of complementary analyses to further validate the performance of the model. The confusion matrix (Figure S3C) demonstrated high ACC, whereas DCA (Figure S4C) confirmed its clinical utility with a favourable net benefit across various thresholds. Moreover, the calibration curve (Figure S5C) showed excellent agreement between the predicted and observed outcomes, and the PR curve (Figure S6C) indicated robust performance, even in the presence of class imbalance. These results collectively validated the LR model as a reliable tool for external datasets. Collectively, these results highlight the robust performance and clinical applicability of the LR model, making it a reliable tool for identifying MASLD in external real-world datasets.

Visualization by SHapley additive explanations (SHAP)

Figure 3B illustrates the importance of the SHAP features for the LR model. The features under scrutiny were arranged in the descending order of their influence on the projected outcomes, as indicated by the mean absolute value of SHAP. The top five pivotal features were weight, height, waist circumference, age, and educational level. The SHAP summary plot further illustrates the impact of these variables on model predictions (Fig. 3A). Notably, higher SHAP values associated with certain features indicated an increased predisposition to developing MASLD. For instance, patients exhibiting aberrant weight were found to be predisposed to a higher incidence of MASLD, compared with those with normal weight.

Interpreting ML models at the patient level

We utilized the SHAP method for individual patient predictions and assess the influence of the LR model on specific patient features. Feature contributions are visually represented by colour coding, where red signifies a positive contribution and blue denotes a negative contribution. For true-positive patients (Figure S7A), the LR model inferred a relatively high probability of developing MASLD, with key features such as weight, height, and educational level exerting a predominant positive influence. Conversely, for true-negative individuals (Figure

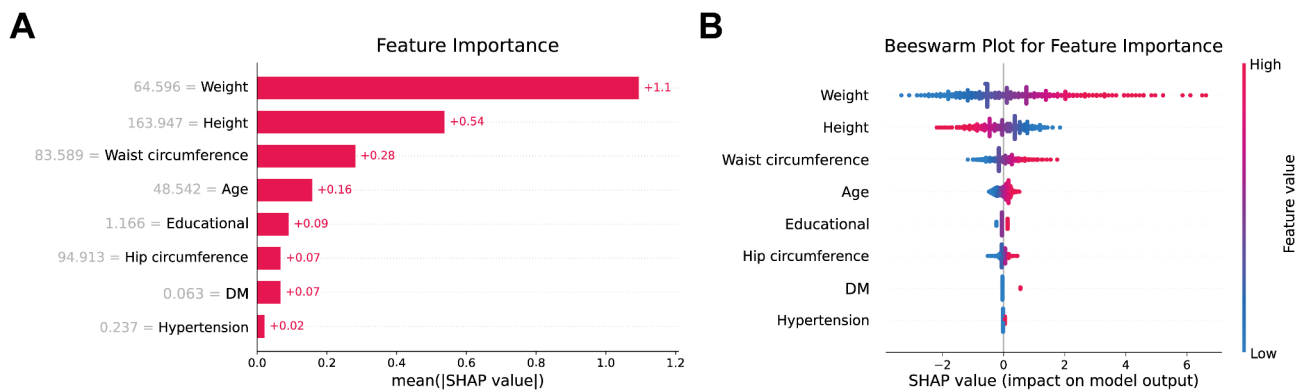


Fig. 3 (A) SHAP summary plot showing the distribution of the SHAP values of each feature; (B) SHAP feature importance shown according to the mean absolute SHAP value of each feature

S7B), the model predicted a lower probability of MASLD, where features such as weight, waist circumference, age, and height exhibited a contrasting contribution pattern.

To further elucidate the interpretability of the LR model, we performed a local interpretable model-agnostic explanation (LIME) analysis to evaluate the contributions of individual features to true-positive (Figure S8A) and true-negative (Figure S8B) predictions. In true-positive predictions, the LR model assigned a high probability (0.87) of developing MASLD (Figure S9A). Features such as height, weight, and waist circumference were identified as the most influential contributors to positive predictions. Negative influences, such as DM and age, had relatively smaller effects on the model output. For true-negative predictions, the model assigned a low probability (0.10) for MASLD (Figure S9B), which was primarily driven by negative contributions from weight and waist circumference. Minimal positive contributions, such as from height and educational level, were insufficient to offset the dominant negative factors. The combined LIME and SHAP analyses consistently highlighted weight and waist circumference as key predictive factors, with the results cross-validating each other and demonstrating the robustness and reliability of the LR model for personalised MASLD risk prediction.

Association between predictors and predicted probability

We utilised restricted cubic splines (RCS) to assess the non-linear relationships between the predictors and the predicted probability of MASLD, adjusting for potential confounders. The analyses demonstrated significant overall associations between the predictors (age, height, weight, BMI, and waist and hip circumference) and the outcomes, as evidenced by p -values of <0.001 for the overall effects across all variables (Figure S10).

To further explore the association between the predictors and MASLD risk, continuous variables identified in the RCS analysis were dichotomised using clinically relevant cutoff values. A forest plot was generated to present

the odds ratios (ORs) and 95% confidence intervals (CIs) for each subgroup (Figure S11). When age was dichotomised using a cut-off of 28 years, individuals aged >28 years showed a significantly higher risk of MASLD, compared with those aged ≤ 28 years (OR: 1.409, 95% CI: 1.174–1.691, $P < 0.001$). Similarly, weight > 56.5 kg was associated with a substantially increased risk of MASLD (OR: 4.887, 95% CI: 4.332–5.512, $P < 0.001$). Height, dichotomized using a cut-off of 158.5 cm, showed an inverse relationship, with individuals taller than the cut-off exhibiting a significantly reduced risk of MASLD (OR, 0.621; 95% CI, 0.551–0.701; $P < 0.001$), compared with those shorter than the cut-off. BMI > 29.4 kg/m² demonstrated the strongest association, conferring an eight-fold increased risk of MASLD (OR, 8.742; 95% CI, 5.541–13.792; $P < 0.001$). In addition, larger waist circumference (> 99 cm) and hip circumference (> 85 cm) were both significantly associated with a higher MASLD risk, with ORs of 3.217 (95% CI: 2.364–4.378, $P < 0.001$) and 1.915 (95% CI: 1.609–2.279, $P < 0.001$), respectively.

Construction and clinical application of an online prediction tool

Building upon the previously established LR model, we developed an online calculator (https://lalalaanjila.shinyapps.io/MASLD_Risk_Prediction_Calculator/) to estimate the probability of MASLD. The tool integrates multiple clinical variables and applies the logit transformation to individual input data, generating a predicted risk probability for MASLD. By offering a user-friendly and precise platform for personalized risk assessment (Figure S12). Compared with conventional scoring systems, this calculator not only improves the predictive accuracy but also streamlines the assessment process, facilitating more efficient individualized treatment strategies. Preliminary validation demonstrated that its performance on the validation set was in line with the original expectations of the model, further supporting its potential for clinical implementation.

Discussion

This study highlighted the application of various ML algorithms to accurately predict the prevalence of MASLD using basic demographic characteristics. The novel approach of utilising 10 distinct algorithms—SVM, MLP, XGBoost, Bagging, KNN, LightGBM, NB, RF, DT, and LR—allowed us to identify the most effective model for MASLD screening.

Each of the tested algorithms demonstrated unique strengths suited to specific application scenarios. LR is widely used due to its simplicity and interpretability; DT provides a transparent decision-making process; RF enhances predictive stability and reduces variance by aggregating multiple DT; SVM excels in handling high-dimensional spaces and capturing complex decision boundaries; KNN performs well in non-parametric classification with well-separated data distributions; MLP, as a neural network approach, captures non-linear patterns effectively; XGBoost and LightGBM demonstrate strong predictive power and computational efficiency, especially in large-scale datasets; Bagging improves model robustness through ensemble learning, and NB is advantageous for probabilistic classification tasks, particularly when feature independence assumptions hold. By comparing the performance metrics of these models, including the ACC, sensitivity, specificity, and AUC, we were able to determine the optimal algorithm for this task. The LR model emerged as the most effective, achieving an impressive ACC of 0.728, a sensitivity of 0.749, and an AUC value of 0.806. This superior performance underscores the capability of LR to balance the classification of positives and negatives, making it a robust tool for early MASLD screening and a crucial reference for future studies.

Furthermore, this study introduced an innovative home-based self-assessment tool for early MASLD detection. This tool, designed to be user-friendly and suitable for domestic use, employs a straightforward questionnaire combined with non-invasive indicators such as age, educational level, height, weight, waist and hip circumference, and history of hypertension and diabetes. Validation in large-scale population studies has demonstrated its efficacy and reliability, offering a practical solution to reduce healthcare costs and enhance patient adherence through early screening and intervention.

The study's utilisation of a large and diverse cohort of 10,007 participants, along with external validation using NHANES data from 2017 to 2023, underscores the robustness and generalisability of our findings. The dataset was systematically partitioned into training, internal testing, and internal validation cohorts, whereas external validations confirmed the performance of the model across various metrics. This extensive cohort not only bolsters the reliability of the results but also enhances the

representativeness of the study, providing a dependable tool for early MASLD screening and serving as a significant reference for future research.

In the landscape of escalating global incidence of MASLD, exacerbated by demographic shifts towards an aging population and changing dietary habits [1], our study addresses a critical gap in early screening and intervention strategies. The current diagnostic and monitoring protocols for MASLD, which are predominantly centralised in specialised medical centres, involve intricate, often costly, and sometimes invasive procedures [35]. This complexity hinders widespread compliance and proactive engagement in MASLD screening, especially when asymptomatic individuals are considered [36]. Therefore, there is an urgent need to develop an accessible and accurate predictive mechanism for MASLD prevalence based solely on fundamental demographic characteristics. Simplifying the diagnostic paradigm to leverage basic baseline features, such as age, educational level, height, weight, and waist and hip circumference, without the need for specialised medical consultations could revolutionise public adherence to liver health screening. As evidenced by our research, this technology has the potential to significantly enhance the ubiquity and compliance of MASLD screening across populations, thereby mitigating the burgeoning burden of liver health issues globally.

Our research, harnessing a selection of robust ML algorithms, demonstrates the feasibility of employing basic demographic and clinical characteristics to predict MASLD in the general population. The significant influences of age, educational level, height, weight, waist and hip circumference, and history of hypertension and diabetes on the development and progression of MASLD were empirically validated. The power of ML to distil these common demographic attributes into actionable insights enables individuals to manage their liver health proactively with minimal intrusion into daily life. Our findings confirmed that ML models can effectively classify the severity of MASLD, with impressive and robust AUC results. We stratified the MASLD group into three categories, offering tailored intervention strategies: (1) individuals with mild or no MASLD, advised to maintain healthy dietary habits and engage in regular physical activity; (2) individuals with moderate MASLD, necessitating more stringent lifestyle and dietary changes, potentially including significant weight management to reduce overall body weight by more than 10%, a crucial factor for improving hepatic fat accumulation [37, 38], along with regular monitoring of liver function and other related parameters; and (3) individuals with severe MASLD, who should seek comprehensive evaluation at specialized liver centres, including liver imaging and biochemical monitoring, with potential recommendations

for medications to regulate blood sugar, fat, and cholesterol levels [39]. This stratification not only streamlines the screening process but also equips healthcare providers with a nuanced approach to effectively manage MASLD at various stages [40].

The validation process revealed the high accuracy and stability of the LR model, which were further corroborated by internal and external validations. The LR model excelled across key performance indicators, consistently outperforming other models in terms of ACC and AUC values, and demonstrating robust generalization capabilities across diverse datasets. These findings provide strong support for the potential of the model for early MASLD detection and intervention, paving the way for future research and clinical applications.

Although our study presents significant advancements in predicting MASLD using basic demographic characteristics, it is important to acknowledge its limitations and scope for further refinement. First, as this was a retrospective study, the selection of outpatient participants may have introduced an inherent bias. Future research should aim for larger-scale prospective studies to validate and replicate these findings. Second, the absence of venous blood sampling for all participants resulted in a lack of comprehensive assessment of relevant liver functions and lipid profiles, which is crucial for a holistic understanding of MASLD [41]. Thirdly, we relied on TE for diagnosing MASLD, as opposed to histopathological biopsy; although practical and non-invasive, TE may not provide the complete picture of the liver condition [15, 42]. This approach is particularly justified in outpatient settings where symptoms are not pronounced and invasive liver biopsies are impractical. Fourth, the potential presence of unrecorded or undiagnosed cases of DM among the screened individuals may have introduced a degree of selection bias, affecting the study results. Finally, the black-box nature of some algorithms can pose challenges in terms of interpretability and transparency, which are critical in medical decision-making. Despite these limitations, the strengths of this study cannot be overlooked. By leveraging simple demographic features, we successfully developed ML models that not only accurately predict MASLD but also stratify patients based on disease severity. This stratification facilitates more personalized intervention and monitoring [43, 44, 45]. Our findings significantly enhance the feasibility and compliance of MASLD screening in the general population, thereby strengthening the public focus on liver health.

Conclusions

This study conclusively demonstrated that ML models, particularly the LR algorithm, effectively utilised a limited number of demographic and readily available clinical characteristics to accurately screen for MASLD in a

broad population. This approach significantly enhances the feasibility and compliance with MASLD screening in general healthcare settings.

Abbreviations

ACC	Accuracy
AUC	Area under the receiver operating characteristic curve
Bagging	Bootstrap aggregating
BMI	Body mass index
CAP	Controlled attenuation parameter
CI	Confidence interval
DCA	Decision curve analysis
DM	Diabetes mellitus
DT	Decision tree
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HOMA-IR	Homeostasis model assessment of insulin resistance
TyG-WC	Triglyceride glucose-waist circumference
KNN	K-nearest neighbours
LASSO	Least absolute shrinkage and selection operator
LightGBM	Light gradient boosting machine
LIME	Local interpretable model-agnostic explanation
LR	Logistic regression
MASLD	Metabolic dysfunction-associated steatotic liver disease
ML	Machine learning
MLP	Multilayer perceptron
NB	Naive bayes
NHANES	National Health and Nutrition Examination Survey
OR	Odds ratio
RCS	Restricted cubic splines
RF	Random forest
SHAP	SHapley additive exPlanations
SVM	Support vector machine
TE	Transient elastography
XGBoost	Extreme gradient boosting

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06387-5>.

Additional File 1: Supplementary tables and figures for machine learning models for predicting metabolic dysfunction-associated steatotic liver disease prevalence using basic demographic and clinical characteristics. This file contains supplementary tables and figures that provide additional details on the study results.

Acknowledgements

The authors acknowledge all the clinical and research staff from the research centers.

Author contributions

GZ: Took the lead in writing the manuscript and performed statistical analyses; YS and ZL: Managed the collection and arrangement of the data; QY and RX: Provided technical support throughout the research process; YX and SG: Engaged in the clinical practices associated with this study; NY and LZ: Reviewed and made significant revisions to the initial draft; XF and RZ: Were responsible for primary data collection; XW, LH and YX: Oversaw the overall direction and planning of the project, and designed the research topic. All authors read and approved the final manuscript.

Funding

The Basic Research Fund, First Affiliated Hospital of Gannan Medical University (QD095).

Data availability

The data included in the study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Review Board of The First Affiliated Hospital of Gannan Medical University (Ganzhou, China). Written informed consent was obtained from all subjects before the study.

Consent for publication

Not applicable.

Competing interests

The authors have declared that no competing interest exists.

Author details

¹The First Clinical Medical College, Gannan Medical University, Ganzhou, Jiangxi Province 341000, China

²Jiangxi Clinical Research Center for Cancer, Department of Oncology, The First Affiliated Hospital, Gannan Medical University, Ganzhou, Jiangxi Province 341000, China

³Department of Rehabilitation Medicine, Affiliated Jinhua Hospital, Zhejiang University School of Medicine, Jinhua, Zhejiang Province 321000, China

⁴Artificial Intelligence Unit, Department of Medical Equipment Management, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

⁵Jiangsu Province Engineering Research Center of Smart Wearable and Rehabilitation Devices, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

⁶Liver Disease Center of Integrated Traditional Chinese and Western Medicine, Department of Radiology, Medical School, Nurturing Center of Jiangsu Province for State Laboratory of AI Imaging & Interventional Radiology, Zhongda Hospital, Southeast University, Southeast University), Nanjing, China

Received: 7 October 2024 / Accepted: 16 March 2025

Published online: 28 March 2025

References

1. Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology* (Baltimore MD). 2023;77(4):1335–47.
2. Stefan N, Cusi K. A global view of the interplay between non-alcoholic fatty liver disease and diabetes. *Lancet Diabetes Endocrinol*. 2022;10(4):284–96.
3. Devasia AG, Ramasamy A, Leo CH. Current therapeutic landscape for metabolic Dysfunction-Associated steatohepatitis. *Int J Mol Sci*. 2025;26(4).
4. Hochreuter MY, Dall M, Treebak JT, Barrès R. MicroRNAs in non-alcoholic fatty liver disease: progress and perspectives. *Mol Metabolism*. 2022;65:101581.
5. Ling Y, Yang YX, Chen YC, Wang JH, Feng DG, Xiang SJ, et al. Newly identified single-nucleotide polymorphism associated with the transition from non-alcoholic fatty liver disease to liver fibrosis: results from a nested case-control study in the UK biobank. *Ann Med*. 2025;57(1):2458201.
6. Tacke F, Puengel T, Loomba R, Friedman SL. An integrated view of anti-inflammatory and antifibrotic targets for the treatment of NASH. *J Hepatol*. 2023;79(2):552–66.
7. Abeysekera KWM, Macpherson I, Glyn-Owen K, McPherson S, Parker R, Harris R, et al. Community pathways for the early detection and risk stratification of chronic liver disease: a narrative systematic review. *Lancet Gastroenterol Hepatol*. 2022;7(8):770–80.
8. Li X, Xiao Y, Chen X, Zhu Y, Du H, Shu J, et al. Machine learning reveals serum glycoproteins as potential biomarkers for the diagnosis of nonalcoholic fatty liver disease (NAFLD). *J Proteome Res*. 2024;23(6):2253–64.
9. Rajewski P, Cieściński J, Rajewski P, Suwała S, Rajewska A, Potasz M. Dietary interventions and physical activity as crucial factors in the prevention and treatment of metabolic Dysfunction-Associated steatotic liver disease. *Biomedicine*. 2025;13(1).
10. Yuan S, Chen J, Li X, Fan R, Arsenault B, Gill D, et al. Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study. *Eur J Epidemiol*. 2022;37(7):723–33.
11. Guagnano MT, D'Ardes D, Ilaria R, Santilli F, Schiavone C, Bucci M et al. Non-Alcoholic fatty liver disease and metabolic syndrome in women: effects of lifestyle modifications. *J Clin Med*. 2022;11(10).
12. Qu B, Li Z. Exploring non-invasive diagnostics for metabolic dysfunction-associated fatty liver disease. *World J Gastroenterol*. 2024;30(28):3447–51.
13. Ramoni D, Liberale L, Montecucco F. Inflammatory biomarkers as cost-effective predictive tools in metabolic dysfunction-associated fatty liver disease. *World J Gastroenterol*. 2024;30(47):5086–91.
14. Long MT, Nouredin M, Lim JK. AGA clinical practice update: diagnosis and management of nonalcoholic fatty liver disease in lean individuals: expert review. *Gastroenterology*. 2022;163(3):764–e741.
15. Yu JH, Lee HA, Kim SU. Noninvasive imaging biomarkers for liver fibrosis in nonalcoholic fatty liver disease: current and future. *Clin Mol Hepatol*. 2023;29(Suppl):S136–49.
16. Choo BP, Goh GBB, Chia SY, Oh HC, Tan NC, Tan JYL, et al. Non-alcoholic fatty liver disease screening in type 2 diabetes mellitus: A cost-effectiveness and price threshold analysis. *Ann Acad Med Singapore*. 2022;51(11):686–94.
17. Wang D, Miao J, Zhang L, Zhang L. Research advances in the diagnosis and treatment of MASLD/MASH. *Ann Med*. 2025;57(1):2445780.
18. Tamaki N, Ajmera V, Loomba R. Non-invasive methods for imaging hepatic steatosis and their clinical importance in NAFLD. *Nat Rev Endocrinol*. 2022;18(1):55–66.
19. Guan X, Chen YC, Xu HX. New horizon of ultrasound for screening and surveillance of non-alcoholic fatty liver disease spectrum. *Eur J Radiol*. 2022;154:110450.
20. Geaney A, O'Reilly P, Maxwell P, James JA, McArt D, Salto-Tellez M. Translation of tissue-based artificial intelligence into clinical practice: from discovery to adoption. *Oncogene*. 2023;42(48):3545–55.
21. Abnoosian K, Farnoosh R, Behzadi MH. Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics*. 2023;24(1):337.
22. Lee J, Westphal M, Vali Y, Boursier J, Petta S, Ostroff R, et al. Machine learning algorithm improves the detection of NASH (NAS-based) and at-risk NASH: A development and validation study. *Hepatology* (Baltimore MD). 2023;78(1):258–71.
23. Ting Sim JZ, Fong QW, Huang W, Tan CH. Machine learning in medicine: what clinicians should know. *Singapore Med J*. 2023;64(2):91–7.
24. Calès P, Canivet CM, Costentin C, Lannes A, Oberti F, Fouchard I et al. A new generation of non-invasive tests of liver fibrosis with improved accuracy in MASLD. *J Hepatol*. 2024.
25. Ben-Assuli O, Jacobi A, Goldman O, Shenhar-Tsarfaty S, Rogowski O, Zeltser D, et al. Stratifying individuals into non-alcoholic fatty liver disease risk levels using time series machine learning models. *J Biomed Inform*. 2022;126:103986.
26. Schophaus S, Creasy KT, Koop PH, Clusmann J, Jaeger J, Punnuru V, et al. Machine learning uncovers manganese as a key nutrient associated with reduced risk of steatotic liver disease. *Liver Int*. 2024;44(10):2807–21.
27. Chen H, Zhang J, Chen X, Luo L, Dong W, Wang Y, et al. Development and validation of machine learning models for MASLD: based on multiple potential screening indicators. *Front Endocrinol (Lausanne)*. 2024;15:1449064.
28. McTeer M, Applegate D, Mesenbrink P, Ratzliff V, Schattenberg JM, Bugianesi E, 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.
29. Nabrdalik K, Kwiendacz H, Irlík K, Hendel M, Drożdż K, Wijata AM, et al. Machine learning identifies metabolic Dysfunction-Associated steatotic liver disease in patients with diabetes mellitus. *J Clin Endocrinol Metab*. 2024;109(8):2029–38.
30. Carrillo-Larco RM, Guzman-Vilca WC, Castillo-Cara M, Alvizuri-Gómez C, Alqahtani S, Garcia-Larsen V. Phenotypes of non-alcoholic fatty liver disease (NAFLD) and all-cause mortality: unsupervised machine learning analysis of NHANES III. *BMJ Open*. 2022;12(11):e067203.
31. Razmpour F, Daryabeygi-Khotbehara R, Soleimani D, Asgharnejad H, Shamsi A, Bajestani GS, et al. Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices. *Sci Rep*. 2023;13(1):4942.
32. Ji W, Xue M, Zhang Y, Yao H, Wang Y. A machine learning based framework to identify and classify Non-alcoholic fatty liver disease in a Large-Scale population. *Front Public Health*. 2022;10:846118.
33. Hu R, Wu B, Wang C, Wu Z, Zhang X, Chen X, et al. Assessment of transient elastography in diagnosing MAFLD and the early effects of sleeve

- gastrectomy on MAFLD among the Chinese population. *Int J Surg (London England)*. 2024;110(4):2044–54.
34. Ajmera V, Cepin S, Tesfai K, Hofflich H, Cadman K, Lopez S, et al. A prospective study on the prevalence of NAFLD, advanced fibrosis, cirrhosis and hepatocellular carcinoma in people with type 2 diabetes. *J Hepatol*. 2023;78(3):471–8.
35. Younossi ZM, Razavi H, Sherman M, Allen AM, Anstee QM, Cusi K, et al. Addressing the High and Rising Global Burden of Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) and Metabolic Dysfunction-Associated Steatohepatitis (MASH): From the Growing Prevalence to Payors' Perspective. *Alimentary pharmacology & therapeutics*; 2025.
36. Dawod S, Brown K. Non-invasive testing in metabolic dysfunction-associated steatotic liver disease. *Front Med (Lausanne)*. 2024;11:1499013.
37. Wei X, Lin B, Huang Y, Yang S, Huang C, Shi L, et al. Effects of Time-Restricted eating on nonalcoholic fatty liver disease: the TREATY-FLD randomized clinical trial. *JAMA Netw Open*. 2023;6(3):e233513.
38. Xue Y, Peng Y, Zhang L, Ba Y, Jin G, Liu G. Effect of different exercise modalities on nonalcoholic fatty liver disease: a systematic review and network meta-analysis. *Sci Rep*. 2024;14(1):6212.
39. Huang DQ, Wong VW, Rinella ME, Boursier J, Lazarus JV, Yki-Järvinen H, et al. Metabolic dysfunction-associated steatotic liver disease in adults. *Nat Reviews Disease Primers*. 2025;11(1):14.
40. Kanwal F, Shubrook JH, Adams LA, Pfothenhauer K, Wai-Sun Wong V, Wright E, et al. Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2021;161(5):1657–69.
41. Deng KQ, Huang X, Lei F, Zhang XJ, Zhang P, She ZG, et al. Role of hepatic lipid species in the progression of nonalcoholic fatty liver disease. *Am J Physiol Cell Physiol*. 2022;323(2):C630–9.
42. Wattacheril JJ, Abdelmalek MF, Lim JK, Sanyal AJ. AGA clinical practice update on the role of noninvasive biomarkers in the evaluation and management of nonalcoholic fatty liver disease. *Expert Rev Gastroenterol*. 2023;165(4):1080–8.
43. Boullion J, Husein A, Agrawal A, Xing D, Hossain MI, Bhuiyan MS et al. Machine Learning-Based biomarker identification for early diagnosis of metabolic Dysfunction-Associated steatotic liver disease. *J Clin Endocrinol Metab*. 2025.
44. Stefanakis K, Mingrone G, George J, Mantzoros CS. Accurate non-invasive detection of MASH with fibrosis F2-F3 using a lightweight machine learning model with minimal clinical and metabolomic variables. *Metab Clin Exp*. 2025;163:156082.
45. Ginter-Matuszewska B, Adamek A, Majchrzak M, Rozplochowski B, Zientarska A, Kowala-Piaskowska A, et al. FibrAlm - The machine learning approach to identify the early stage of liver fibrosis and steatosis. *Int J Med Informatics*. 2025;197:105837.

Publisher's note

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