EJPMR, 2026, 13(8), 644-652 **Research Article** **Impact Factor: 8.158**

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

ISSN (O): 2394-3211
ISSN (P): 3051-2573
Coden USA: EJPMAG

AN EXPLAINABLE MACHINE LEARNING APPROACH FOR NON-INVASIVE PREDICTION OF METABOLIC DYSFUNCTION-ASSOCIATED STEATOTIC LIVER DISEASE USING ROUTINE LABORATORY BIOMARKERS

Ipek Balikci Cicek[1]*, Zeynep Kucukakcali[1]

Inonu University Faculty of Medicine, Department of Biostatistics and Medical Informatics, 44280, Malatya, Turkey.

***Corresponding Author: Ipek Balikci Cicek**

Inonu University Faculty of Medicine, Department of Biostatistics and Medical Informatics, 44280, Malatya, Turkey.

DOI: https://doi.org/10.5281/zenodo.21886356

How to cite this Article: Ipek Balikci Cicek[1]*, Zeynep Kucukakcali[1]. (2026). An Explainable Machine Learning Approach For Non-Invasive Prediction Of Metabolic Dysfunction-Associated Steatotic Liver Disease Using Routine Laboratory Biomarkers. European Journal of Pharmaceutical and Medical Research, 13(8), 644–652.

Article Received on 16/07/2026 **Article Revised on 06/08/2026** **Article Published on 10/08/2026**

ABSTRACT

Background: Metabolic dysfunction-associated steatotic liver disease (MASLD) is the leading cause of chronic liver disease worldwide, yet population-level screening relies largely on imaging unavailable in primary care. We aimed to develop and explain machine learning models that predict MASLD using routine laboratory biomarkers alone. **Methods:** We analyzed 8,008 adults from the National Health and Nutrition Examination Survey (NHANES) 2017-March 2020 cycle with valid vibration-controlled transient elastography (VCTE) measurements. MASLD was defined per the 2023 AASLD/EASL/ALEH consensus nomenclature as hepatic steatosis (controlled attenuation parameter [CAP] ≥248 dB/m, with ≥285 dB/m as a sensitivity threshold) plus at least one of five cardiometabolic risk criteria, after excluding participants with significant alcohol consumption. To avoid circularity, predictors were restricted to variables not used in outcome construction**:** age, sex, race/ethnicity, income-poverty ratio, ALT, AST, GGT, the AST/ALT ratio, albumin, creatinine, uric acid, and C-reactive protein. Five classifiers (logistic regression, random forest, XGBoost, LightGBM, CatBoost) were compared using nested cross-validation (5-fold outer, 3-fold inner) for hyperparameter tuning and unbiased performance estimation. Model behavior was explained using SHAP (global importance, dependence, and interaction values) and LIME (instance-level explanations). As a methodological contribution, we quantified (1) per-patient agreement between SHAP and LIME explanations (Spearman correlation) and (2) cross-model stability of SHAP-based feature rankings (Kendall's tau) among the four tree-based models. **Results:** MASLD prevalence was 56.0% (CAP≥248 dB/m) and 36.4% (CAP≥285 dB/m). CatBoost achieved the highest nested cross-validation performance (ROC-AUC 0.786 ± 0.007), closely followed by random forest, XGBoost and LightGBM (all ROC-AUC 0.78-0.79 on the held-out test set); logistic regression trailed at 0.752, indicating a modest but genuine benefit from non-linear modeling. Results were consistent, and slightly stronger, under the more stringent CAP≥285 dB/m definition (CatBoost ROC-AUC 0.798). The AST/ALT ratio was the dominant predictor across all five models, showing a sharp transition around a ratio of 1 that mirrors known hepatological patterns (ratio <1 associated with simple steatosis, >1 with more advanced disease), followed by age, CRP, uric acid, and GGT. Feature-ranking agreement among the four tree-based models was high (mean Kendall's tau = 0.86), and SHAP-LIME explanation agreement was high for most patients (mean Spearman ρ = 0.84) but diverged for a small subgroup, warranting caution in individual-level decision support. **Conclusions:** Routine, widely available laboratory biomarkers, combined with explainable machine learning, can identify adults with likely MASLD with fair-to-good discrimination and clinically coherent explanations, offering a low-cost pre-screening tool to prioritize patients for confirmatory imaging.

KEYWORDS: MASLD; NAFLD; explainable AI; SHAP; LIME; NHANES; machine learning; liver steatosis.

1. INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD), the condition formerly known as non-alcoholic fatty liver disease (NAFLD), affects an estimated one in three adults worldwide and is projected to become the leading indication for liver transplantation in the coming decade.[1] An interim 2020 proposal first renamed the disease metabolic dysfunction-associated fatty liver disease (MAFLD) to emphasize its metabolic underpinnings,[2] and the 2023 multi-society consensus refined this further to MASLD, requiring hepatic steatosis in the presence of at least one cardiometabolic risk factor and the exclusion of excessive alcohol use or competing etiologies.[3]

Diagnosis traditionally relies on imaging, including ultrasound, magnetic resonance imaging-based proton density fat fraction, or vibration-controlled transient elastography (VCTE, e.g., FibroScan),[4] which is not universally accessible in primary care and is impractical for population-level screening. Machine learning models built on routine laboratory panels offer an inexpensive, scalable pre-screening alternative, but their clinical adoption is limited by a lack of interpretability, a concern that is particularly acute for high-stakes clinical decisions.[5] Explainable AI (XAI) techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) address this gap by attributing individual predictions to specific input features, but the two methods are rarely cross-validated against one another, and the stability of explanations across different model architectures is seldom reported, both of which matter if explanations are to inform clinical trust.

Most prior NHANES-based machine learning studies of fatty liver disease used surrogate, formula-based steatosis indices, such as the Fatty Liver Index[6] or Hepatic Steatosis Index,[7] as the outcome, and frequently included the same anthropometric or metabolic variables in both the outcome definition and the predictor set, a source of circularity that inflates apparent performance and distorts explanation-based feature rankings. The present study addresses these gaps by (1) using an objective, VCTE-measured hepatic steatosis outcome rather than a formula-based proxy, (2) defining MASLD per the current 2023 consensus nomenclature, (3) deliberately excluding outcome-defining variables from the predictor set to avoid tautological associations, (4) benchmarking five classifiers under nested cross-validation, and (5) introducing two explanation-focused analyses, namely SHAP-LIME agreement and cross-model explanation stability, as quantitative measures of explanation trustworthiness.

2. METHODS

2.1 Data Source and Study Population

Data were obtained from the National Health and Nutrition Examination Survey (NHANES) 2017-March 2020 pre-pandemic cycle, a nationally representative, cross-sectional survey conducted by the U.S. Centers for Disease Control and Prevention (CDC). NHANES combines standardized interviews, physical examinations, and laboratory testing; all data are publicly available and de-identified, and secondary analysis of these public-use files does not constitute human subjects research requiring additional institutional review. Demographic, body-measurement, laboratory, questionnaire, and liver ultrasound transient elastography files were downloaded directly from the CDC NHANES data repository and merged on the unique participant identifier (SEQN).

Eligible participants were adults (≥18 years) with a valid VCTE-derived controlled attenuation parameter (CAP) measurement. Of 9,693 adults in the merged dataset, 8,317 (85.8%) had a valid CAP measurement. Participants with significant alcohol consumption (estimated weekly intake ≥21 drinks for men or ≥14 drinks for women, derived from self-reported drinking frequency [ALQ121] multiplied by typical quantity per drinking day [ALQ130]) were excluded to isolate a metabolic, rather than alcohol-related, phenotype, removing 309 participants (3.7%). The final analytic cohort comprised 8,008 adults.

2.2 Outcome Definition

MASLD was defined according to the 2023 AASLD/EASL/ALEH multi-society consensus nomenclature, requiring both of the following[3]

(1) Hepatic steatosis, defined by CAP ≥248 dB/m (primary threshold, corresponding to the S≥S1 cut-off, graded relative to the histological reference standard for steatosis [9], identified in the largest individual-patient-data meta-analysis of CAP-biopsy correlation studies) [10] or, as a sensitivity analysis, the more stringent CAP ≥285 dB/m, approximating the "severe steatosis" threshold (>280 dB/m) used in prior NHANES-based VCTE studies.[11]

(2) At least one of five cardiometabolic risk criteria, adapted from the harmonized metabolic syndrome definition[8]: (a) body mass index ≥25 kg/m² or elevated waist circumference (>102 cm men, >88 cm women); (b) glycemic dysfunction (HbA1c ≥5.7%, fasting glucose ≥100 mg/dL, physician-diagnosed diabetes, or antidiabetic medication use); (c) elevated blood pressure (≥130/85 mmHg, averaged across up to three oscillometric readings, or antihypertensive medication use); (d) triglycerides ≥150 mg/dL; (e) low HDL cholesterol (≤40 mg/dL men, ≤50 mg/dL women).

Two binary outcomes were constructed ($MASLD_{248}$ and $MASLD_{285}$) to test the robustness of downstream findings to the steatosis threshold. $MASLD_{248}$ was designated the primary outcome; $MASLD_{285}$ served as a sensitivity analysis.

2.3 Predictors

Twelve predictors were selected: age, sex, race/ethnicity, family income-to-poverty ratio, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), the AST/ALT ratio (De Ritis ratio), albumin, creatinine, uric acid, and high-sensitivity C-reactive protein (CRP). Variables used to construct the cardiometabolic component of the outcome, namely body mass index, waist circumference, HbA1c, fasting glucose, diabetes status, blood pressure, triglycerides, and HDL cholesterol, were deliberately excluded from the predictor set. Including them would create circularity: a model could trivially recover part of its own label definition rather than learning a genuine association, inflating performance metrics and distorting SHAP-based feature attributions. Missing predictor values (range 0-13.8% across variables; see Table 1) were imputed using the training-set median within each cross-validation fold, a simple and widely used approach for handling missing data at these moderate missingness levels.[12]

2.4 Model Development

Model development and reporting followed the general principles of the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guideline [13]. Data were split into training (80%, n=6,406) and held-out test (20%, n=1,602) sets, stratified by outcome. Five classifiers were evaluated: logistic regression (with standardized predictors, included as a linear baseline), random forest,[14] XGBoost,[15] LightGBM,[16] and CatBoost.[17] Hyperparameters were tuned using nested cross-validation, which provides a less optimistically biased performance estimate than single-layer cross-validation with hyperparameter selection: [18,19] an outer 5-fold stratified loop provided an unbiased estimate of generalization performance, while an inner 3-fold grid search selected hyperparameters within each outer training fold, preventing information leakage from hyperparameter selection into performance estimation. Final models used for interpretation and test-set evaluation were separately tuned via 3-fold grid search on the full training set. Discrimination was assessed with the area under the receiver operating characteristic curve (ROC-AUC), F1-score, precision, and recall; calibration, an often under-reported but clinically important property of risk prediction models,[21] was assessed with the Brier score[20] and calibration curves (10 bins).

2.5 Explainability Analysis

Global and local explanations were generated using SHAP (SHapley Additive exPlanations): [22] TreeExplainer was used for the four tree-based models (exact Shapley values), and a model-agnostic permutation-based explainer was used for logistic regression. For the best-performing model (by nested cross-validation ROC-AUC), we additionally computed SHAP dependence plots and exact pairwise SHAP interaction values[23] (computed on a random subsample of 200 test patients for computational tractability). LIME (Local Interpretable Model-agnostic Explanations) [24] was used to generate instance-level explanations for three representative patients (a high-confidence positive case, a high-confidence negative case, and a case near the decision boundary).

Two explanation-focused analyses were performed as methodological contributions of this study. First, SHAP-LIME agreement was quantified for 100 randomly sampled test patients by computing the Spearman rank correlation between each patient's SHAP feature-contribution vector and their corresponding LIME feature-weight vector; a higher correlation indicates that the two independent explanation methods agree on which features drove that patient's prediction. Second, cross-model explanation stability was quantified among the four tree-based models by computing Kendall's tau rank correlation between each pair of models' global (mean absolute SHAP value) feature rankings; logistic regression was excluded from this comparison because its explanation mechanism (linear coefficients via a model-agnostic explainer) is not directly comparable to Shapley values derived from tree structure.

2.6 Software

All analyses were performed in Python 3 using scikit-learn, XGBoost, LightGBM, CatBoost, SHAP, and LIME. Random seeds were fixed (seed=42) throughout for reproducibility.

3. RESULTS

3.1 Cohort Characteristics

Among 8,008 eligible adults, $MASLD_{248}$ prevalence was 56.0% (4,486/8,008) and $MASLD_{285}$ prevalence was 36.4% (2,913/8,008). These figures are consistent with previously published NHANES 2017-2020 VCTE-based MASLD prevalence estimates (~57%) [25,26,27], supporting the validity of our outcome construction. Missingness among predictors ranged from 0% (age, sex, race/ethnicity) to 13.8% (income-poverty ratio); liver enzymes and other laboratory predictors had 7.0-7.4% missingness (Table 1).

Table 1: Predictor variables and missingness.

Variable	Description	Missing (%)
Age	Age in years	0.0
Sex	Male / Female	0.0
Race/Ethnicity	NHANES race/ethnicity category	0.0
Income-Poverty Ratio	Family income-to-poverty ratio	13.8

ALT	Alanine aminotransferase (U/L)	7.0
AST	Aspartate aminotransferase (U/L)	7.4
GGT	Gamma-glutamyl transferase (U/L)	7.0
AST/ALT Ratio	Derived (De Ritis ratio)	7.4
Albumin	Serum albumin (g/dL)	7.0
Creatinine	Serum creatinine (mg/dL)	7.0
Uric Acid	Serum uric acid (mg/dL)	7.0
CRP	High-sensitivity C-reactive protein (mg/L)	7.2

3.2 Model Performance

Table 2 summarizes nested cross-validation and held-out test performance for $MASLD_{248}$. CatBoost achieved the highest nested cross-validation ROC-AUC (0.786 ± 0.007), followed closely by random forest (0.783 ± 0.005), XGBoost (0.781 ± 0.005), and LightGBM (0.781 ± 0.005); the four tree-based models were within 0.005 ROC-AUC of one another, indicating that performance was not strongly sensitive to the specific gradient-boosting or ensemble algorithm chosen. Logistic regression achieved a lower but still reasonable ROC-AUC (0.759 ± 0.007), suggesting that non-linear modeling provided a modest, genuine improvement over a linear baseline rather than reflecting overfitting artifacts: nested cross-validation and held-out test performance were closely aligned for every model (e.g., CatBoost: 0.786 nested vs. 0.779 test), indicating minimal optimism bias. Brier scores (0.186-0.201) indicated reasonably well-calibrated probability estimates across all models, consistent with the calibration curves in Figure 1.

Table 2: Model performance for $MASLD_{248}$ prediction (nested cross-validation and held-out test set).

Model	Nested CV ROC-AUC	Test ROC-AUC	Test F1	Test Precision	Test Recall	Test Brier
Logistic Regression	0.759 ± 0.007	0.752	0.737	0.700	0.778	0.201
Random Forest	0.783 ± 0.005	0.769	0.748	0.699	0.805	0.192
XGBoost	0.781 ± 0.005	0.778	0.757	0.712	0.809	0.188
LightGBM	0.781 ± 0.005	0.779	0.753	0.707	0.806	0.188
CatBoost	0.786 ± 0.007	0.779	0.751	0.708	0.798	0.188

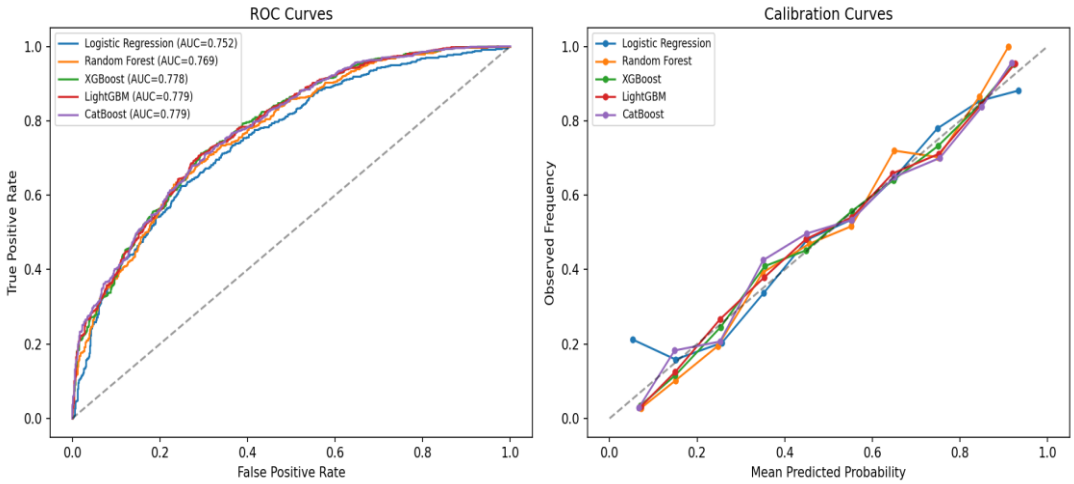

Figure 1: ROC curves (left) and calibration curves (right) for all five models on the held-out test set.

3.3 Sensitivity Analysis

Under the more stringent $MASLD_{285}$ definition, test-set ROC-AUC was similar to or slightly higher than under $MASLD_{248}$ for every model (logistic regression 0.760, random forest 0.786, XGBoost 0.789, LightGBM 0.790, CatBoost 0.798), indicating that model performance was robust to, and if anything strengthened by, the choice of steatosis threshold (Table 3).

Table 3: Sensitivity analysis: test ROC-AUC under the CAP≥285 dB/m ($MASLD_{285}$) threshold.

Model	$MASLD_{285}$ Test ROC-AUC
Logistic Regression	0.760
Random Forest	0.786
XGBoost	0.789
LightGBM	0.790
CatBoost	0.798

3.4 Global Feature Importance

SHAP global importance rankings were highly consistent across all five models (Figure 2 shows the best-performing model, CatBoost; corresponding plots for all models are provided in the Supplement). The AST/ALT ratio was the single most important predictor in every model, followed by age, CRP, uric acid, and GGT. Sex and income-poverty ratio contributed least across all models.

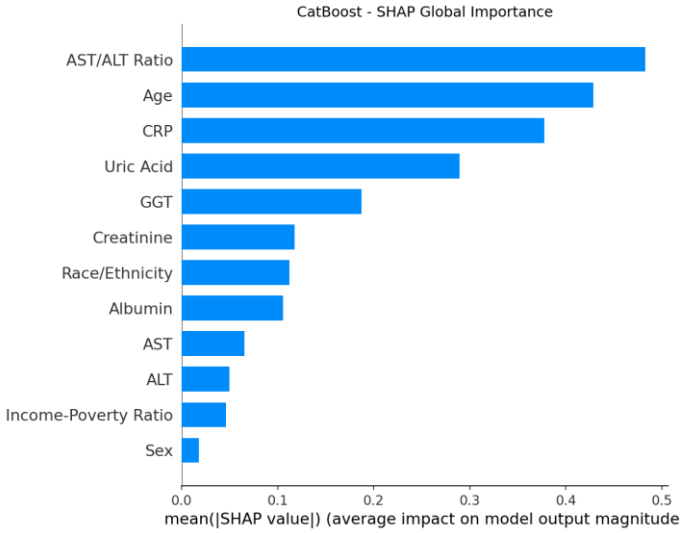

Figure 2: SHAP global feature importance for CatBoost (best-performing model by nested cross-validation ROC-AUC).

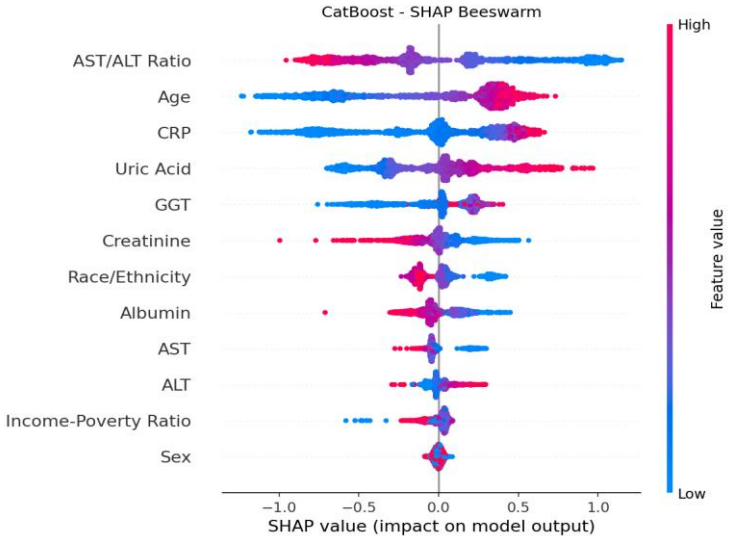

Figure 3: SHAP beeswarm plot for CatBoost, showing the direction and distribution of each feature's contribution across the test set.

3.5 Dependence and Interaction Effects

The SHAP dependence plot for the AST/ALT ratio (Figure 4) revealed a sharp, sigmoid-shaped transition around a ratio of approximately 1.0: values below 1 contributed positively to the MASLD prediction, while values above 1 contributed strongly negatively. This pattern is consistent with established hepatological knowledge: ALT typically exceeds AST in simple hepatic steatosis, whereas the ratio reverses (AST > ALT) with progression toward fibrosis or alcohol-related injury, providing external, literature-based validation that the model captured a clinically coherent, rather than spurious, pattern. The SHAP interaction summary (Figure 5) showed that the AST/ALT ratio's contribution varied only modestly with other features, indicating a largely independent (additive) effect, whereas age showed a visible interaction pattern consistent with cardiometabolic risk accumulating with age.

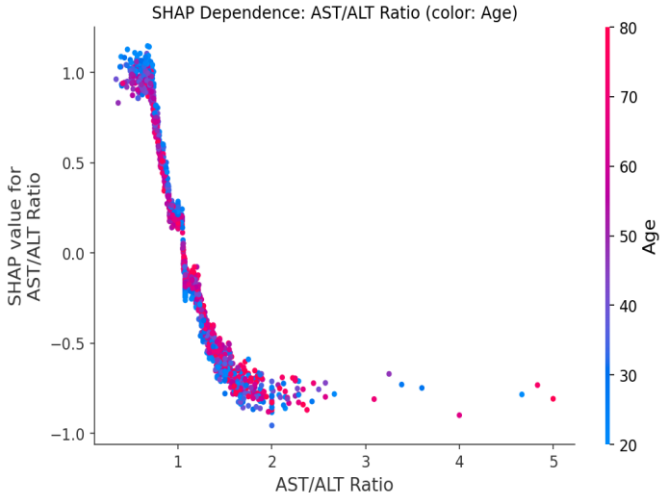

Figure 4: SHAP dependence plot for the AST/ALT ratio, colored by age, for the CatBoost model.

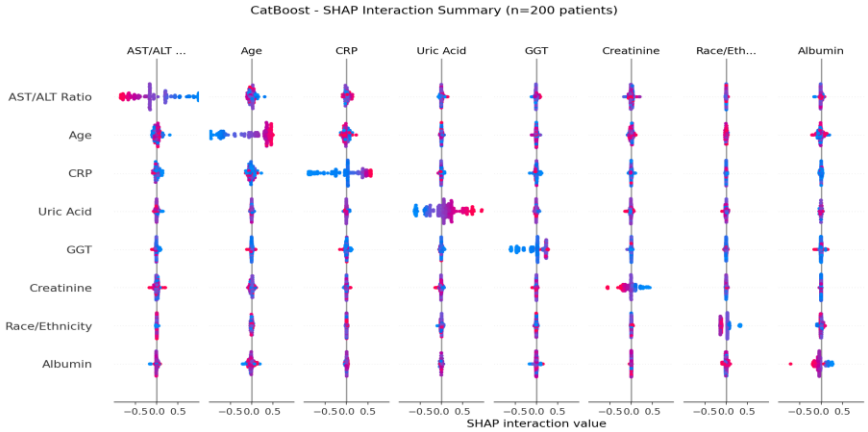

Figure 5: SHAP interaction value summary for the top features (CatBoost, n=200 randomly sampled test patients).

3.6 SHAP-LIME Explanation Agreement

Across 100 randomly sampled test patients, the per-patient Spearman correlation between SHAP and LIME feature attributions averaged 0.84, with the majority of patients (>80%) showing correlations above 0.85, indicating strong agreement between the two independent explanation methods for most individuals (Figure 6). A small subset of patients (approximately 5-6%) showed markedly lower agreement ($\rho < 0.3$), suggesting that for a minority of cases the two methods emphasize different features, a pattern that warrants caution when using either method alone to justify an individual clinical decision, and that we flag as a direction for future work (e.g., characterizing what distinguishes these low-agreement cases).

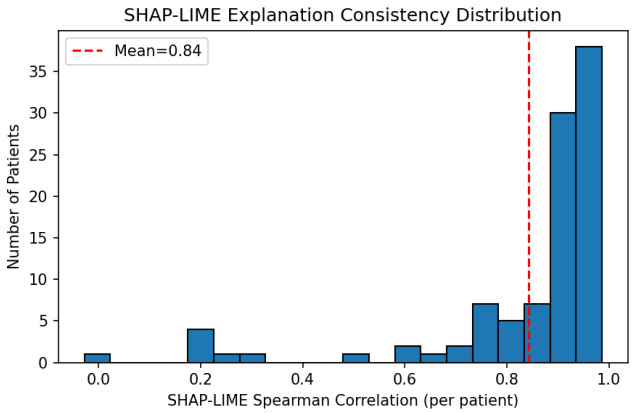

Figure 6: Distribution of per-patient SHAP-LIME Spearman correlation (n=100 test patients, CatBoost model).

3.7 Cross-Model Explanation Stability

Kendall's tau rank correlations between the four tree-based models' global SHAP feature rankings were uniformly high (Table 4; mean tau = 0.86, range 0.73-1.00), with XGBoost and LightGBM showing essentially perfect agreement (tau = 1.00) and both showing near-perfect agreement with CatBoost (tau = 0.97). Random forest showed somewhat lower, but still substantial, agreement with the gradient-boosting models (tau = 0.73-0.76). All pairwise correlations were statistically significant ($p < 0.001$). This high level of cross-model agreement indicates that the identified predictors of MASLD are a robust, model-independent signal rather than an artifact of any single algorithm's inductive bias.

Table 4: Cross-model explanation stability: pairwise Kendall's tau between SHAP feature rankings (tree-based models only).

Model 1	Model 2	Kendall's tau	p-value
Random Forest	XGBoost	0.727	<0.001
Random Forest	LightGBM	0.727	<0.001
Random Forest	CatBoost	0.758	<0.001
XGBoost	LightGBM	1.000	<0.001
XGBoost	CatBoost	0.970	<0.001
LightGBM	CatBoost	0.970	<0.001

4. DISCUSSION

This study developed and rigorously explained machine learning models that predict MASLD, defined per current 2023 consensus criteria using an objective VCTE-based outcome, from twelve routine demographic and laboratory variables. Four tree-based models achieved similar discrimination (test ROC-AUC 0.77-0.78), modestly but consistently outperforming a linear baseline (0.75), with results that were robust to an alternative, more stringent steatosis threshold. These findings are consistent with, and in several respects methodologically stronger than, prior NHANES-based fatty liver prediction studies, most of which relied on formula-based steatosis proxies and did not explicitly guard against circularity between outcome definition and predictor selection.

The dominance of the AST/ALT ratio, and its literature-consistent directionality (low ratio favoring an MASLD prediction, high ratio disfavoring it), provides a form of external validation that is often missing from black-box risk models: the pattern the model learned matches an independently established hepatological mechanism, since the AST/ALT (De Ritis) ratio has long been recognized as rising with progressive liver injury and fibrosis rather than simple steatosis [29], rather than reflecting an artifact of the training data. Uric acid, the fourth-ranked predictor across all models, has similarly been independently associated with hepatic steatosis and fibrosis in prior NHANES VCTE-based analyses,[28] and GGT, another top-ranked predictor, has independently been proposed as a marker of the oxidative stress that accompanies hepatic fat accumulation,[30] further supporting the biological plausibility of the model's explanations. Race/ethnicity ranked seventh in importance across models, a finding consistent with well-documented disparities in NAFLD/MASLD prevalence and severity across US racial and ethnic groups.[31] This kind of mechanistic consistency is precisely the value proposition of explainable AI in a clinical context: it allows clinicians to audit whether a model's reasoning aligns with domain knowledge before trusting its output.

Two explanation-focused analyses were introduced as methodological contributions. First, cross-model explanation stability was high among tree-based models (mean Kendall's tau = 0.86), suggesting that the top predictors identified here are unlikely to be an artifact of model choice; this is an important consideration given that most published XAI-in-healthcare studies report explanations from a single model without testing whether a different, similarly performing model would highlight the same features. Second, SHAP-LIME agreement was high on average (mean $\rho = 0.84$) but showed a non-trivial tail of low-agreement cases, indicating that explanation reliability is not uniform across patients and should ideally be assessed, not merely assumed, before an explanation is used to support an individual clinical decision.

4.1 Limitations

Several limitations warrant discussion. First, this is a cross-sectional analysis; associations should not be interpreted causally. Second, NHANES complex survey design (sampling weights, strata, and clusters) was not applied in this analysis; unweighted estimates are appropriate for internal predictive modeling but do not directly generalize to national population-level prevalence without weighting. Third, alcohol consumption was self-reported and the weekly-drinks estimate used for exclusion is an approximation; residual misclassification of alcohol-related liver disease as MASLD cannot be excluded. Fourth, CAP thresholds for steatosis grading are not universally standardized across VCTE probe types and populations; [10,32] we addressed this partially via a two-threshold sensitivity analysis, but external, biopsy-confirmed validation was not available. Fifth, all data derive from a single national survey (United States); generalizability to other populations and healthcare systems requires external validation. Finally, no external, temporally or geographically distinct validation cohort was available; nested cross-validation

provides an honest internal performance estimate but is not a substitute for external validation.

5. CONCLUSION

Explainable machine learning models built on twelve routine demographic and laboratory variables, deliberately excluding variables used to define the outcome, predicted MASLD with fair-to-good discrimination (ROC-AUC 0.77-0.79) and clinically coherent, cross-model-stable explanations. The AST/ALT ratio emerged as the dominant, mechanistically plausible predictor. High agreement between SHAP and LIME explanations for most, but not all, patients underscores both the promise and the current limits of explainable AI for individual-level clinical decision support. These models offer a low-cost, imaging-free pre-screening approach that could help prioritize patients for confirmatory elastography or biopsy, particularly in resource-limited settings.

REFERENCES

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease: meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology, 2016; 64(1): 73-84.
2. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. J Hepatol., 2020; 73(1): 202-209.
3. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al.; NAFLD Nomenclature consensus group. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol., 2023; 79(6): 1542-1556.
4. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology, 2019; 156(5): 1264-1281.
5. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell, 2019; 1(5): 206-215.
6. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol., 2006; 6: 33.
7. Lee JH, Kim D, Kim HJ, Lee CH, Yang JI, Kim W, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis., 2010; 42(7): 503-508.
8. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation, 2009; 120(16): 1640-1645.
9. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology, 2005; 41(6): 1313-1321.
10. Karlas T, Petroff D, Sasso M, Fan JG, Mi YQ, de Lédinghen V, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. J Hepatol., 2017; 66(5): 1022-1030.
11. Trivedi HD, Niezen S, Jiang ZG, Tapper EB. Severe hepatic steatosis by controlled attenuation parameter predicts quality of life independent of fibrosis. Dig Dis Sci., 2022; 67(8): 4215-4222.
12. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol., 2006; 59(10): 1087-1091.
13. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ., 2015; 350: g7594.
14. Breiman L. Random forests. Mach Learn, 2001; 45(1): 5-32.
15. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.
16. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems (NeurIPS), 30. 2017: 3146-3154.
17. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems (NeurIPS) 31. 2018: 6638-6648.
18. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 2006; 7: 91.
19. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res., 2010; 11: 2079-2107.
20. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev., 1950; 78(1): 1-3.
21. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med., 2019; 17(1): 230.
22. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), 30. 2017: 4765-4774.

23. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell, 2020; 2(1): 56-67.
24. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1135-1144.
25. Luo N, Zhang X, Huang J, Chen H, Tang H. Prevalence of steatotic liver disease and associated fibrosis in the United States: results from NHANES 2017-March 2020. J Hepatol., 2024; 80(2): e70-e71.
26. Kalligeros M, Vassilopoulos A, Vassilopoulos S, Victor DW, Mylonakis E, Noureddin M. Prevalence of steatotic liver disease (MASLD, MetALD, and ALD) in the United States: NHANES 2017-2020. Clin Gastroenterol Hepatol., 2024; 22(6): 1330-1332.e4.
27. Zhang X, Heredia NI, Balakrishnan M, Thrift AP. Prevalence and factors associated with NAFLD detected by vibration controlled transient elastography among US adults: results from NHANES 2017-2018. PLoS One, 2021; 16(6): e0252164.
28. Duan H, Zhang R, Chen X, Yu G, Song C, Jiang Y, et al. Associations of uric acid with liver steatosis and fibrosis applying vibration controlled transient elastography in the United States: a nationwide cross-section study. Front Endocrinol (Lausanne), 2022; 13: 930224.
29. Botros M, Sikaris KA. The de Ritis ratio: the test of time. Clin Biochem Rev., 2013; 34(3): 117-130.
30. Lee DH, Blomhoff R, Jacobs DR Jr. Is serum gamma glutamyltransferase a marker of oxidative stress? Free Radic Res., 2004; 38(6): 535-539.
31. Rich NE, Oji S, Mufti AR, Browning JD, Parikh ND, Odewole M, et al. Racial and ethnic disparities in nonalcoholic fatty liver disease prevalence, severity, and outcomes in the United States: a systematic review and meta-analysis. Clin Gastroenterol Hepatol., 2018; 16(2): 198-210.e2.
32. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology, 2019; 156(6): 1717-1730.