Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Interpretable Machine Learning Predicts Hepatic Fat Response to a Monounsaturated Fat-Enriched Diet in Steatotic Liver Disease

Alejandro Miron Jabalera

mironalejandro@g.ucla.edu

University of California, Los Angeles

Jieping Yang

University of California, Los Angeles

Hila Zelicha

University of California, Los Angeles

William Hsu

University of California, Los Angeles

Paivi Pajukanta

University of California, Los Angeles

Bui Alex

University of California, Los Angeles

Zhaoping Li

University of California, Los Angeles

Tzung Hsiai

University of California, Los Angeles

Article

Keywords: artificial intelligence, and machine-learning (AI/ML), metabolic dysfunction-associated steatotic liver disease (MASLD), hepatic fat fraction, Insulin resistance, visceral adiposity

Posted Date: June 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9794767/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Interpretable Machine Learning Predicts Hepatic Fat Response to a Monounsaturated Fat-Enriched Diet in Steatotic Liver Disease

AM Jabalera [1], J Yang [2,5], H Zelicha [2,5], W Hsu [3], P Pajukanta [4], A Bui [1,3], Z Li [5], T Hsiai [1,2,6]

[1] *Department of Bioengineering, University of California, Los Angeles, CA, USA*

[2] *Department of Medicine, Division of Cardiology, University of California, Los Angeles, CA, USA*

[3] *Department of Radiological Sciences, Medical Imaging Informatics, University of California, Los Angeles, CA, USA*

[4] *Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*

[5] *Center for Human Nutrition, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*

[6] *VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA*

Corresponding Author: ***Alejandro Miron Jabalera – AM Jabalera (mironalejandro@g.ucla.edu)***

ABSTRACT

Background: Metabolic dysfunction-associated steatotic liver disease (MASLD) varies in response to dietary monounsaturated fatty acids (MUFA). We conducted a secondary analysis of the Habitual Diet vs Avocado Trial (HAT; NCT03528031; registered 17 May 2018; n=790) to test whether machine learning using routinely collected clinical variables can identify MUFA-responsive phenotypes. Avocado versus habitual diet did not significantly change hepatic-fat fraction (ΔHFF +0.48 percentage points, 95% CI −0.24 to 1.22; p=0.20) or visceral adipose tissue volume after FDR correction. Prespecified interaction tests showed sex- and glycemic-status-specific responses (FDR-q ≤ 0.05). MASLD classification reached AUROC 0.83 ± 0.03 (XGBoost) but its expected calibration error (ECE = 0.089) exceeded the prespecified ≤0.05 target, indicating that probability outputs require post-hoc Platt or isotonic recalibration before any thresholded clinical use. Two related HFF prediction tasks were evaluated: (i) the primary trial-cohort endpoint of prespecified ≥5 percentage-point absolute hepatic fat change versus no such change (either direction; n=790; 261 positive), for which discrimination was 0.67 ± 0.03 with baseline biomarkers and rose to ~0.75 ± 0.03 when baseline MASLD status was carried forward as an input feature; and (ii) a translation-focused reduction-only endpoint (≥5 pp absolute decrease; listwise complete predictors, n=790; 74 positives), summarized in Figure 7, for which XGBoost achieved held-out AUROC=0.792. Gains under (i) reflect sequential staging plus baseline-outcome correlation; confusion matrices must use endpoint-specific true-class row totals. Reporting follows TRIPOD+AI. Internally validated; external validation is required before clinical decision-support deployment.

Keywords: artificial intelligence, and machine-learning (AI/ML), metabolic dysfunction-associated steatotic liver disease (MASLD), hepatic fat fraction, Insulin resistance, visceral adiposity.

Introduction

Artificial intelligence (AI) and machine learning (ML) approaches are increasingly used in biomedical research to model complex, high-dimensional clinical data that are not adequately captured by conventional statistical methods [1,2]. In conditions characterized by metabolic heterogeneity and interacting cardiometabolic risk factors, ML methods can complement traditional regression analyses by identifying nonlinear relationships, higher-order interactions, and clinically relevant subgroups. These methodologies are particularly valuable in clinical trials and nutritional interventions, where metabolic responses to diet have been demonstrated in precision-nutrition studies, supporting the rationale for predictive approaches to dietary intervention [7].

Metabolic dysfunction-associated steatotic liver disease (MASLD) affects nearly one-third of adults worldwide and is now defined under the 2023 multisociety steatotic liver disease nomenclature [3,4]. It is rapidly emerging as a leading driver of cardiometabolic morbidity and mortality. MASLD is closely linked to cardiometabolic morbidity and cardiovascular disease risk, and contemporary guidelines continue to emphasize lifestyle intervention as a foundation of management [5,6]. However, responses to dietary interventions vary considerably across individuals, reflecting differences in insulin resistance, adiposity distribution, metabolic phenotype, and diabetic status. As a result, population-level dietary recommendations often fail to capture clinically meaningful heterogeneity in treatment response.

Monounsaturated fatty acids (MUFA), including those provided by avocados, have been associated with favorable cardiometabolic profiles in observational and interventional studies [8]. The Habitual Diet and Avocado Trial (HAT) provides a well-characterized randomized cohort for evaluating whether neutral average effects of avocado supplementation on visceral adiposity and hepatic fat may conceal clinically meaningful response heterogeneity [9,10]. In the parent HAT trial, one avocado per day for 26 weeks did not significantly reduce visceral adipose tissue volume compared with habitual diet, and no significant group differences were observed for hepatic fat fraction, hsCRP, or metabolic-syndrome components [10]. Modest nominal reductions in total and LDL cholesterol were observed, but the overall intervention effect on adiposity and hepatic fat was neutral [10]. Subsequent HAT analyses showed improved HEI-2015 diet quality and changes in gut-

microbiome diversity with avocado supplementation, supporting intervention fidelity and biological dietary exposure despite neutral average adiposity effects [11,12]. Therefore, the present secondary analysis focuses on MASLD phenotyping, treatment-response heterogeneity, and machine-learning prediction rather than re-estimating the parent trial's primary efficacy analysis.

Recent advances in computational modeling provide new opportunities to integrate multidimensional clinical, metabolic, and anthropometric data (e.g., body mass index, waist circumference) to improve phenotypic classification and predict treatment response. ML methods trained on routinely available clinical variables can identify complex patterns that may inform individualized risk stratification and therapeutic targeting. For these reasons, these AI methods can complement, not replace, traditional statistical models by providing predictive frameworks that identify responders and nonresponders within heterogeneous trial populations.

We re-analyzed the HAT trial (NCT03528031; n=790) to test three prespecified hypotheses: (i) baseline MASLD status (HFF >5%) can be discriminated from routinely available clinical variables with AUROC ≥0.80; (ii) the pipeline-defined hepatic fat decrease endpoint used for the primary responder dashboard (HFF_Responder; percentile-based decrease bin) and, separately, the translation endpoint of ≥5 percentage-point absolute reduction in HFF (wk26 − wk0 ≤ −5 pp), can each be discriminated above chance using baseline clinical variables; and (iii) prespecified interaction tests will reveal sex- and glycemic-status-specific responses to the intervention. Outputs are intended to inform a future deployable decision-support tool for hepatology, primary-care, and dietitian use.

Methods

Study design and participants

The HAT Study (ClinicalTrials.gov ID: NCT03528031) was a multicenter, randomized, controlled trial of adults with abdominal obesity conducted across multiple US academic medical centers [9,10]. Eligible participants were ≥25 years, with a waist circumference ≥35 inches for women and ≥40 inches for men. Primary exclusion criteria included other chronic liver diseases, decompensated cirrhosis, significant cardiovascular instability, pregnancy, and MRI contraindications. Participants were randomly allocated a 1:1 ratio to an avocado-supplemented diet or a habitual diet without avocado for 26 weeks [9,10]. Comprehensive details on inclusion and exclusion criteria, randomization procedures, and study protocols are provided in the Supplementary Methods section [9,10]. The Habitual Diet vs Avocado Trial (HAT; ClinicalTrials.gov ID:NCT03528031, registered: 17 May 2018) screened 1,008 adults with abdominal obesity, of whom 923 had complete baseline magnetic-resonance proton-density fat-fraction (MRI-PDFF) measurements [9,10]. For the present predictive analysis, we restricted the cohort to participants with paired baseline (Week 0) and end-of-intervention (Week 26) measurements. Of 965 participants with at least one complete clinical visit in the analytic source file, 175 contributed only a single visit (early termination, missed Week-26 visit, or sample failure) and were excluded; 790 participants (396 avocado, 394 habitual; 73.2% female, 26.8% male) had complete paired data and constitute the analytic cohort. We excluded 14 participants for missing electronic health record (EHR) data, 27 for lipidomic data, 41 for biomarker data, 60 for social determinants of health (SDoH) data, and 70 for MRI-PDFF data. Another 162 participants were excluded for missing repeated measures. As a result, 790 participants were included in the final analysis. Among the 790 participants, approximately 55% had HFF greater than 5% and were classified as having MASLD.

Glycemic status was classified at baseline using fasting plasma glucose (FPG) measured after an overnight fast, applying the American Diabetes Association 2024 criteria [13]: normoglycemic (FPG <100 mg/dL), prediabetes (FPG 100-125 mg/dL), and diabetes (FPG ≥126 mg/dL). Participants on glucose-lowering medication at screening were classified as having diabetes regardless of measured FPG. The resulting distribution in the 790-participant analytic cohort was 374 normoglycemic (47.3%), 332 prediabetes (42.0%), and 84 diabetes (10.6%); these classes were balanced across intervention arms (χ^2 $p = 0.557$; all standardized mean differences $|SMD| \leq 0.08$).

For the purposes of this predictive analysis, we operationalize MASLD as MRI-PDFF-quantified hepatic fat fraction greater than 5%, the imaging criterion of the 2023 multi-society MASLD nomenclature [3] consistent with contemporary MASLD nomenclature and clinical practice guidance [3,6]. The cardiometabolic-risk-factor component of that definition (≥1 of: BMI ≥25 kg/m², elevated waist circumference, fasting glucose ≥100 mg/dL, blood pressure ≥130/85 mmHg, plasma triglycerides ≥150 mg/dL, or HDL <40/<50 mg/dL) was not separately enforced because all participants in the parent trial were enrolled with abdominal obesity, which alone satisfies the cardiometabolic-risk-factor criterion.

Targeted metabolomics, targeted lipidomic, and erythrocyte (red-blood-cell) membrane fatty acids were assayed at baseline (Week 0) and end of intervention (Week 26) for the majority of analytic-cohort participants; the corresponding paired-sample size for each platform is reported in the relevant Results subsection. Untargeted metabolomics and untargeted lipidomic were available on a smaller subset and contribute only to the exploratory `with untargeted` predictive-analysis branch.

We further filtered 166 variables for initial feature selection and excluded 29 variables with >50% missing data. Consequently, we retained 137 variables in the final training set (Figure 1). Analyses of 26-week changes used pre-post

difference scores calculated within each participant, with between-group comparisons performed on these difference scores using bootstrapped independent-samples tests. This analysis was conducted and reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD+AI) guidelines for clinical prediction models using machine learning methods [14]. A completed TRIPOD+AI checklist is provided as Supplementary Table S5. For retained variables, missing data were handled via median/mode imputation fitted on the training folds. As a limitation, we did not perform multiple imputation or formal sensitivity analyses comparing imputation strategies.

Figure1. Study Flow and Cohort Selection

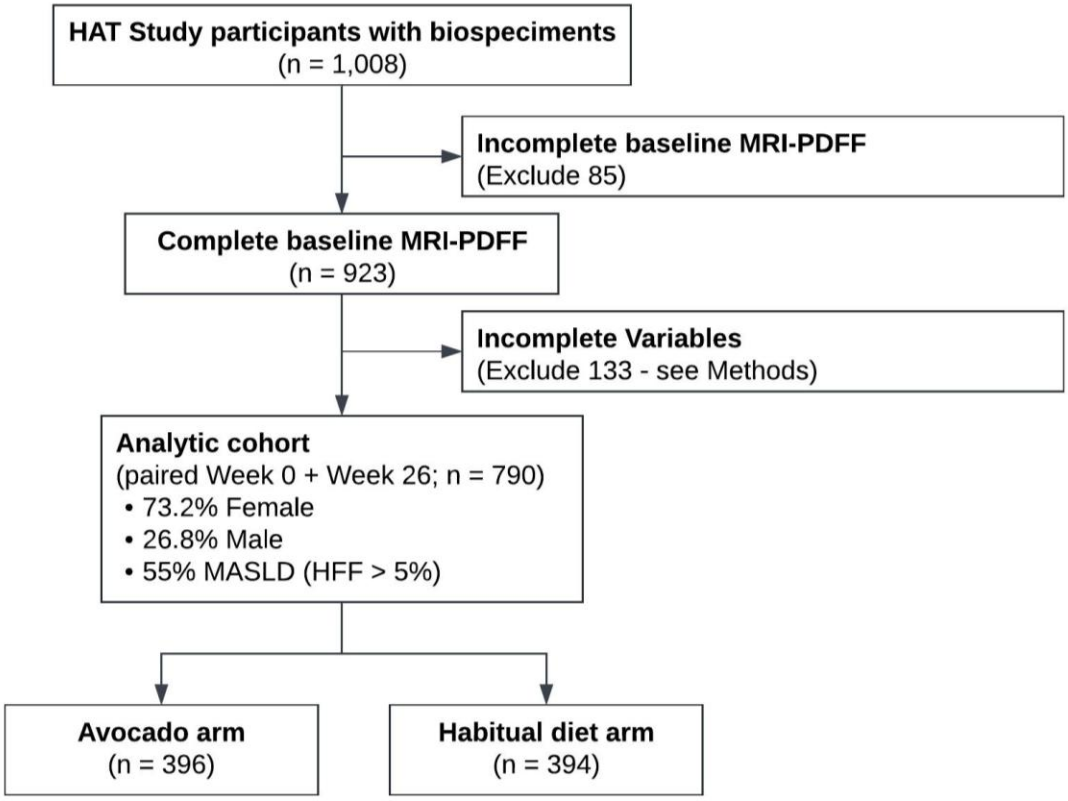


Figure 1. Study Flow and Cohort Selection. CONSORT diagram depicting screening, eligibility assessment, exclusions, and final analytic sample for the Habitual Diet and Avocado Trial (HAT). Of 1,008 initially screened participants (1,936 samples), exclusions were sequentially applied for missing electronic health record (EHR) variables (n=14), lipidomic data (n=27), biomarker data (n=41), social determinants of health (SDoH) variables (n=60), and MRI-proton density fat fraction (MRI-PDFF) data (n=70). An additional 162 participants without repeat 26-week measurements were excluded from longitudinal analyses. The final analytic cohort comprised 790 randomized participants with complete baseline and follow-up data included in the primary outcomes analysis.

Data Integration

To evaluate the specific contribution of baseline hepatic steatosis status, we conducted a sequential feature branching analysis comparing models with and without binary MASLD status (HFF>5% yes) and continuous HFF. Furthermore, we assessed the value of high-dimensional metabolic profiling by comparing models using the targeted clinical/lipidomic feature set against models that additionally incorporated untargeted lipidomic variables. Platform-specific sample sizes and baseline missingness are reported in Supplementary Table S4.

Imaging and Subgroup Analyses

Hepatic fat fraction (HFF) and visceral adipose tissue volume (VAT) were quantified at baseline and at week 26 utilizing standardized MRI-PDFF. HFF was calculated as the mean PDFF across multiple regions of interest within the right hepatic lobe. VAT was determined through semi-automated segmentation of visceral fat at the L4-L5 vertebral level, with quality control conducted by expert readers who were blinded to the participants' randomization arm [9,10].

The primary imaging outcomes for this analysis were changes in HFF and VAT from baseline to week 26. We focused on pre-specified subgroups defined by sex (female, male) and glycemic status (without diabetes, pre-diabetic, diabetic). Treatment effects within each subgroup were summarized as mean differences with 95% confidence intervals (CI) and visualized in forest plots. Overall avocado-versus-habitual-diet effects on visceral adiposity, hepatic fat fraction, hsCRP,

and cardiometabolic risk factors have been reported previously from the parent HAT trial. The present secondary analysis therefore focused on MASLD-stratified heterogeneity, hepatic-fat response phenotyping, and machine-learning prediction rather than re-estimating the primary efficacy analysis of avocado supplementation.

Correlation and Regression Analyses

For baseline-change and dietary-clinical associations examined within prespecified subgroups, we estimated Pearson and Spearman correlations separately within each subgroup stratum. We used paired bootstrap resampling to construct confidence intervals for the correlation coefficient (pivotal intervals by default; percentile bootstrap if selected) and permutation-based two-sided p-values. Covariate-adjusted partial Pearson correlations-both variables residualized on age, BMI, sex, and study-site indicators, were summarized with approximate Fisher-z confidence intervals and are reported in parallel with simple correlations. Exploratory scatterplots displayed an ordinary least squares fit and, in strata with sufficient sample size ($n \geq 20$), linear quantile regression lines at the 25th, 50th, and 75th percentiles [16]; OLS, Huber, and quantile slopes were also tabulated for descriptive comparison [16,17]. Separately, treatment-by-subgroup interaction models were fit by multivariable linear regression; heteroscedasticity-consistent sandwich standard errors (HC3) were applied to those regression coefficients when the Breusch-Pagan test indicated heteroscedasticity ($p < 0.05$) [18,19]. Correlation results are reported as r or ρ (95% CI), two-sided p, and n within each subgroup.

Multiplicity correction was applied within the prespecified subgroup-interaction family using Benjamini-Hochberg FDR over four tests (sex×treatment for ΔHFF and ΔVAT; glycemic-status×treatment for ΔHFF and ΔVAT) [20]. The adjusted-treatment-effect FDR family covers the 12 prespecified clinical outcomes. All other tests (correlation grids, decision-curve thresholds, fatty-acid-by-arm tests, threshold sweeps) are exploratory and reported uncorrected; we treat associations from these analyses as hypothesis-generating, with confirmatory inference deferred to external-validation cohorts.

Predictive modeling: feature sets and outcomes

To perform the subgroup analyses, we developed AI-enabled predictive models utilizing two related outcome sets. Initially, within a broader EHR-derived cohort that includes individuals with and without MASLD, we modeled baseline MASLD status (present versus absent). Subsequently, within the trial cohort, we modeled (A) the primary binary endpoint HFF_Responder carried in the prepared machine-learning matrix (MRI-PDFF decrease below the pipeline's data-driven lower bound for balanced change categorization; n = 790 with 261 positives, 33.0%), and (B) a secondary translation endpoint of ≥5 percentage-point reduction only (wk26 - wk0 ≤ -5 pp) among participants with listwise complete predictors (n = 790; 74 positives, 9.4%). Because MRI-PDFF is a validated quantitative marker of hepatic steatosis and change in MRI-PDFF has been linked with histologic response in NASH, we evaluated a pragmatic secondary translation endpoint of ≥5 percentage-point absolute HFF reduction [21-25]. We also acknowledge that guidelines and meta-analyses more commonly emphasize ≥30% relative MRI-PDFF decline as a histologic response threshold [6,25]. The primary ML dashboard therefore reflects the pipeline's percentile-anchored decrease label, whereas reduction-only figures use the fixed ≥5 pp clinical anchor. To evaluate a sequential diagnostic-to-therapeutic pipeline, we assessed whether incorporating baseline MASLD status or continuous baseline HFF improved performance for each endpoint.

Candidate predictors included participants' demographics, anthropometrics, clinical laboratory values, lipidomic measures, and selected social determinants of health (SDoH). These included age, sex, BMI, weight, waist circumference, blood pressure, fasting glucose and insulin levels, insulin resistance (HOMA-IR), standard lipid profiles (LDL, HDL, triglycerides, VLDL cholesterol), high-sensitivity C-reactive protein (hs-CRP), and RBC fatty acids as palmitoleic acid (MFA16:1n-7) and trans-palmitoleic acid (TFA16:1n-7T). SDoH included education, living status, and food access indicators, where such data were available [26] (Table 1). While no formal sample-size calculation was performed a priori for this secondary machine learning analysis, event-per-variable heuristics depend on the endpoint: the primary HFF_Responder model (n = 790; 261 positive events across ~29 predictors) yields ~9 events per variable, below the common ≥10 events-per-variable heuristic [27] but with a substantially enriched positive class versus the reduction anchor. The reduction-only endpoint (74 positives in n = 790) is far sparser and should be interpreted as hypothesis-generating with respect to EPV. Separately, the MASLD classifier uses n=790 (binary model predicting baseline HFF >5% from ten bootstrap-selected baseline features; n=100 resamples, ≥60% selection frequency; see Methods). To maintain clarity, the primary figures present these subgroup effects and predictive modeling results, with computational methods using bootstrapping. A comprehensive codebook defining all continuous and categorical predictors used in the final models, alongside their baseline missingness proportions, is detailed in Supplementary Table S4.

Model development and explainability

To mitigate overfitting and identify stable predictors, we employed a bootstrap-based variable selection method [28]. From 100 bootstrap resamples of the training dataset, random forest models were fitted, feature importances were calculated, predictors were ranked, and the rankings were aggregated to determine a core set of approximately ten features

that were both informative and stable across resamples. Highly collinear predictors were excluded based on pairwise correlation analysis.

Using this reduced feature set, we trained three principal algorithms: L2-regularized logistic regression, random forest [29], and XGBoost [30]. Hyperparameters were tuned with stratified cross-validation, and models were evaluated with bootstrap validation on held-out data. For each algorithm and outcome, we estimated the area under the receiver operating characteristic curve (AUROC) and computed accuracy, precision, recall, F1-score, and specificity after dichotomizing predicted probabilities using the Youden-optimal threshold [32]. The manuscript primarily focuses on bootstrapped AUROC, calibration, and threshold-dependent held-out performance metrics.

Internal validation strategy

Models were developed on a stratified 70/30 train/test partition (single random seed = 42; class-stratified). Within the training fold we performed three nested operations: (i) bootstrap variable selection - 100 bootstrap resamples of the training set, random-forest median-importance ranking, top-10 features carried forward, with collinearity pruning; (ii) hyperparameter tuning by RandomizedSearchCV - 5-fold StratifiedKFold for the MASLD-classification pipeline (48 search iterations) and 5-fold × 2-repeat RepeatedStratifiedKFold for the change/responder pipelines (50 search iterations), scoring=ROC-AUC; (iii) final-fit on the full training fold using the selected hyperparameters. Performance on the held-out 30% test set was summarized by 100-iteration stratified-bootstrap resampling of the test set (yielding mean AUROC ± SD and 95% CI) together with point estimates from a fixed seed. Because the train/test partition itself was a single random split, we additionally performed a resample-then-split sensitivity analysis using the dual_bootstrap_comparison.py workflow; the corresponding AUROC distributions and source data are archived with the public analysis code repository. This design constitutes an internal-validation estimate with k-fold-tuned hyperparameters and bootstrap-quantified test-set sampling variance; external validation in an independent MASLD cohort remains required before clinical-decision-support deployment.

Calibration

Calibration was assessed using the expected calibration error (ECE: 10 equal-width probability bins) and the Brier score [33,34]. Calibration plots and numerical ECE and Brier score values for all ML models are provided in Supplementary Figure S9 and Supplementary Table S3. The pre-specified ECE threshold for acceptable calibration of the primary MASLD classification model was ECE < 0.05. To enhance interpretability, we utilized Shapley Additive exPlanations (SHAP) to assess each predictor's marginal contribution to individual-level model predictions [35]. We aggregated SHAP values across bootstrap iterations to identify predictors that consistently increased or decreased the predicted probability of MASLD or HFF response. A multi-panel figure delineates variable importance and SHAP-derived effects for the most clinically pertinent predictors. Hyperparameters for all classification models were selected via randomized search (RandomizedSearchCV, scikit-learn [31]) over pre-specified parameter distributions with 48 to 50 search iterations (48 for the MASLD classification pipeline and 50 for the longitudinal-change pipeline) and 5-fold stratified cross-validation. Detailed specifications of the machine learning algorithms, including hyperparameter search spaces, regularization methods to handle class imbalance, and cross-validation strategies, are provided in Supplementary Table S3. Class imbalance was addressed through balanced sample weights (sklearn compute_sample_weight='balanced'); no synthetic over-sampling was applied to the primary outcomes. Model performance was evaluated on a single stratified held-out test set (30% of the analytic sample, allocated prior to any model training) and summarized across 100 bootstrap replicates resampling the test set with replacement, yielding mean AUROC ± SD and 95% confidence intervals. This design constitutes a held-out test set with inner stratified cross-validation for hyperparameter tuning; it is not a fully nested cross-validation scheme, and reported performance should be interpreted as an internal validation estimate from a single partition. Calibration was assessed using expected calibration error (ECE; 10 equal-width probability bins); we pre-specified that the primary MASLD classification model should achieve ECE < 0.05 (numerical results reported in Results and Supplementary Table S3). All predictive modeling and statistical analyses were performed in Python version 3.11 using scikit-learn (v1.3), xgboost (v2.0), and shap (v0.42) within a standardized computing environment. This prediction-model analysis was secondary to the parent trial; the prespecified statistical analysis plan and the completed TRIPOD+AI checklist are provided as Supplementary Table S5.

Results

Study population

Data from 790 participants from the Habitual Diet versus Avocado Trial (HAT) were used in this analysis, including 396 assigned to avocado supplementation and 394 assigned to the habitual diet control group. Participant enrollment and randomization occurred between June 27, 2018, and March 4, 2020, with final 6-month follow-up imaging and clinical assessments completed by October 2020. The study cohort included approximately 73.2% of female, and 26.8% male. At

baseline, more than half of (55%) participants met criteria for metabolic dysfunction-associated steatotic liver disease (MASLD), defined as hepatic fat fraction (HFF) ≥5%. Baseline HFF varied widely across the 790-participant analytic cohort, ranging from 0.07% to 54.11%, with a mean of 10.45% and median of 6.28%, indicating broad variation in hepatic steatosis burden. Subgroup-specific baseline MASLD prevalence is shown in Supplementary Figure S1, and baseline HFF, VAT, and hsCRP distributions stratified by sex and glycemic status are shown in Supplementary Figure S2.

Baseline demographic, metabolic, and anthropometric characteristics were assessed across the cohort, comparing patients with and without baseline MASLD (Table 1). Continuous variables are reported as medians with 95% bootstrap confidence intervals, and categorical variables as counts and percentages. Between-group comparisons were performed using the Wilcoxon rank-sum test for continuous variables and χ^2 or Fisher's exact tests for categorical variables, as appropriate.

Table 1. Bivariate Analysis of Baseline Characteristics by MASLD Status

Common name	MASLD (N=438)	CONTROL (N=352)	Difference (95% CI)	P (FDR, bootstrap)
Sex				
Female (n=578)	282	296		**2.27×10^{-6}**
Male (n=212)	156	56		
Race / Ethnicity				
White	317	242		0.1272
Black	50	61		
Asian	25	14		
Other	24	22		
Hispanic	22	13		
Intervention				
Avocado Diet	219	177		0.9938
Habitual Diet	219	175		
Time point				
Week 0	438	352		
Week 26	438	352		
EHR				
Age (years)	51.2 (49.9, 52.5)	49.6 (48.1, 51.1)	1.6 (-0.36, 3.6)	0.1202
Weight (lb)	113.3 (109.3, 117.1)	107.7 (102.7, 112.5)	5.5 (-0.76, 11.8)	0.1011
Height (ft)	5.3 (5.2, 5.3)	5.1 (5.0, 5.2)	0.17 (0.05, 0.3)	**0.0122**
BMI (kg/m²)	33.9 (33.4, 34.4)	30.8 (30.4, 31.3)	3.1 (2.4, 3.8)	**2.27×10^{-6}**
Waist circumference (in)	5.1 (5.0, 5.2)	4.1 (4.0, 4.2)	1.0 (0.85, 1.2)	**2.27×10^{-6}**
Diastolic BP, ln(mmHg)	4.4 (4.3, 4.4)	4.3 (4.3, 4.3)	0.04 (0.02, 0.06)	**2.31×10^{-5}**
Systolic Blood Pressure (mmHg)	125.9 (124.4, 127.4)	119.2 (117.5, 120.8)	6.8 (4.5, 9.0)	**2.27×10^{-6}**
Beats per minute (BPM)	70.1 (69.2, 71.1)	68.3 (67.3, 69.3)	1.8 (0.44, 3.2)	**0.0153**
Biomarkers				
Insulin (µIU/mL)	22.7 (20.8, 24.4)	11.8 (10.8, 12.7)	10.9 (8.8, 12.9)	**2.27×10^{-6}**
hsCRP (mg/L)	7.3 (6.4, 8.1)	4.6 (4.0, 5.1)	2.8 (1.7, 3.8)	**4.17×10^{-6}**
Glucose (mg/dL)	112.9 (109.8, 115.8)	98.6 (97.1, 99.9)	14.4 (11.0, 17.7)	**2.27×10^{-6}**
Cholesterol (0.01×mg/dL)	192.5 (188.6, 196.3)	184.7 (180.7, 188.7)	7.7 (2.2, 13.3)	**0.0118**
Triglycerides (mg/dL)	151.5 (141.8, 160.6)	97.5 (92.1, 102.6)	54.0 (43.2, 64.8)	**2.27×10^{-6}**
HDL (mmol/L)	3.9 (3.9, 3.9)	4.0 (4.0, 4.0)	-0.11 (-0.14, -0.08)	**2.27×10^{-6}**
LDL (mg/dL)	114.2 (111.0, 117.5)	110.1 (106.6, 113.5)	4.1 (-0.66, 8.9)	**0.1108**
VLDL (mg/dL)	30.3 (28.4, 32.1)	19.5 (18.4, 20.5)	10.8 (8.6, 13.0)	**2.27×10^{-6}**
HOMA-IR	6.8 (6.0, 7.4)	2.9 (2.7, 3.2)	3.8 (3.1, 4.6)	**2.27×10^{-6}**
Glaecemic Status	0.82 (0.75, 0.88)	0.41 (0.35, 0.46)	0.41 (0.32, 0.5)	**2.27×10^{-6}**
SDOH				
Education	3.6 (3.5, 3.7)	3.9 (3.8, 4.0)	-0.32 (-0.51, -0.13)	**0.0023**
Living Situation	1	1		
Smoking Status				
Current Smoker: Yes	8	13		0.0739
Current Smoker: No	428 (98)	333 (95)		
History of Smoking: Yes	94	79		0.9168
History of Smoking: No	342 (78)	272 (77)		
Physical Activity				
Regular Activity: Yes	276	250		0.0514
Regular Activity: No	161 (37)	100 (28)		

Table 1. Bivariate Analysis of Baseline Characteristics by MASLD Status. Comparison of demographic, metabolic, anthropometric, and clinical characteristics at baseline between participants with MASLD (HFF ≥5%; n=438) and without MASLD (HFF <5%; n=352). Continuous variables reported as medians (95% bootstrap CI); categorical variables as counts and percentages. P-values from Wilcoxon rank-sum test (continuous) or χ^2/Fisher's exact test (categorical). FDR-adjusted p-values (Benjamini-Hochberg) provided. Boldface indicates $p<0.05$.

Primary outcomes

The parent HAT trial previously reported the overall avocado-versus-habitual-diet effects on MRI-derived visceral adiposity and secondary cardiometabolic outcomes [10]. In that primary trial analysis, one avocado per day for 26 weeks did not significantly reduce visceral adipose tissue volume compared with habitual diet, and no significant group differences were observed for hepatic fat fraction, hsCRP, or metabolic-syndrome components [10]. Modest nominal reductions in total and LDL cholesterol were observed, but the overall intervention effect on adiposity and hepatic fat was neutral [10].

Therefore, the present analysis does not reinterpret the parent trial as a primary efficacy study; instead, it asks whether baseline MASLD status, cardiometabolic phenotype, and machine-learning models can identify response heterogeneity that was not apparent from the average treatment effect.

To complement the previously reported parent-trial efficacy analysis, we retained Table 2 as a descriptive longitudinal phenotype analysis stratified by baseline MASLD status. This analysis compares 26-week within-participant changes in clinical and biomarker variables between participants with and without baseline MASLD, pooled across intervention arms, and is not intended to estimate the causal effect of avocado supplementation. Intervention assignment was balanced between baseline MASLD and non-MASLD groups, supporting interpretation of Table 2 as a MASLD-status-stratified trajectory analysis rather than an arm-comparison analysis. Most anthropometric, blood-pressure, glycemic, inflammatory, and insulin-resistance changes did not differ significantly by baseline MASLD status after FDR correction. The clearest between-group difference was observed for total cholesterol change, which decreased more among participants with baseline MASLD than among controls. Triglyceride and VLDL changes showed directionally greater reductions in the MASLD group but did not meet the FDR-adjusted significance threshold. These findings support the interpretation that baseline MASLD status captured a metabolically distinct subgroup with modestly different lipid trajectories over follow-up, while the parent HAT trial remains the appropriate source for the overall avocado-versus-habitual-diet efficacy results [10].

Table 2. Bivariate Analysis of 26-Week Changes by Baseline MASLD Status

Common name	MASLD (N=438)	CONTROL (N=352)	Difference (95% CI)	P (FDR, bootstrap)
Sex				
Female (n=578)	282	296		**2.40×10^{-5}**
Male (n=212)	156	56		
Race / Ethnicity				
White	317	242		0.351
Black	50	61		
Asian	25	14		
Other	24	22		
Hispanic	22	13		
Intervention				
Avocado Diet	219	177		1.000
Habitual Diet	219	175		
Change (Week 0 to Week 26)				
N with change data	438	352		
EHR				
Δ Weight (lb)	8.6 (5.5, 11.4)	8.0 (5.1, 10.6)	0.61 (-3.5, 4.7)	1.000
Δ Height (ft)	0 (0, 0)	0 (0, 0)	0 (—, —)	1.000
Δ BMI (kg/m²)	0 (0, 0)	0 (0, 0)	0 (—, —)	1.000
Δ Waist circumference (in)	-0.18 (-0.6, 0.25)	0.11 (-0.36, 0.58)	-0.29 (-0.93, 0.35)	0.820
Δ Diastolic BP, ln(mmHg)	-0.5 (-2.0, 0.63)	-0.89 (-1.7, -0.07)	0.38 (-1.2, 1.9)	1.000
Δ Systolic Blood Pressure (mmHg)	-1.2 (-2.3, -0.09)	-0.86 (-2.1, 0.35)	-0.35 (-2.0, 1.3)	1.000
Δ Beats per minute (BPM)	0.69 (-0.05, 1.4)	0.36 (-0.41, 1.1)	0.34 (-0.73, 1.4)	1.000
Biomarkers				
Δ Insulin (µIU/mL)	-0.28 (-1.6, 1.1)	0.03 (-0.59, 0.65)	-0.31 (-1.8, 1.2)	1.000
Δ hsCRP (mg/L)	0.09 (-0.82, 0.92)	1.1 (0.02, 1.9)	-0.99 (-2.3, 0.31)	0.355
Δ Glucose (mg/dL)	1.3 (-0.63, 3.3)	0.78 (-0.24, 1.8)	0.57 (-1.6, 2.8)	1.000
Δ Cholesterol (0.01×mg/dL)	-7.5 (-9.7, -5.3)	-2.2 (-4.9, 0.51)	-5.3 (-8.8, -1.8)	**0.031**
Δ Triglycerides (mg/dL)	-6.3 (-12.3, -0.18)	4.5 (-0.32, 9.5)	-10.8 (-18.6, -3.0)	0.055
Δ HDL (mmol/L)	-0.57 (-1.2, 0.06)	0.24 (-0.55, 1.0)	-0.81 (-1.8, 0.2)	0.351
Δ LDL (mg/dL)	-5.8 (-7.8, -3.9)	-3.5 (-6.0, -1.0)	-2.3 (-5.4, 0.87)	0.357
Δ VLDL (mg/dL)	-1.3 (-2.5, -0.04)	0.91 (-0.06, 1.9)	-2.2 (-3.7, -0.6)	0.055
Δ HOMA-IR	0.1 (-0.53, 0.66)	0.02 (-0.16, 0.2)	0.08 (-0.54, 0.7)	1.000
Δ Diabetes	0.03 (-0.03, 0.08)	0.02 (-0.03, 0.08)	0 (-0.07, 0.08)	1.000

Table 2. Bivariate Analysis of 26-Week Changes by Baseline MASLD Status. Comparison of 26-week changes in clinical variables between participants with and without MASLD at baseline, pooled across intervention arms. Continuous variables reported as medians (95% bootstrap CI). Boldface indicates $p<0.05$.

Heterogeneity of treatment response

Despite the overall neutral treatment effects, prespecified interaction analyses revealed significant differences in response among key subgroups. After adjusting for multiple comparisons using the Benjamini-Hochberg false discovery rate method, a significant treatment-by-sex interaction was identified for VAT (p interaction=0.02; FDR-adjusted p=0.04). Among female participants, avocado supplementation was associated with a greater reduction in visceral adiposity than in the control group (mean difference 0.052 L; 95% CI 0.010 to 0.096; p=0.013), whereas no significant effect was observed in male participants (Figures 2D and 2E).

Glycemic status also influenced treatment response. Significant interactions between treatment and glycemic status were observed for both VAT (p interaction=0.03; FDR-adjusted p=0.045) and HFF (p interaction=0.04; FDR-adjusted p=0.048). Participants without diabetes showed reductions in visceral adiposity (mean difference 0.072 L; 95% CI 0.013 to

0.134; p=0.017), whereas individuals with prediabetes had a paradoxical increase in hepatic fat (mean difference 1.29%; 95% CI 0.11 to 2.60; p=0.039). No significant effects were detected among participants with established diabetes (Figure 2F and 2G).

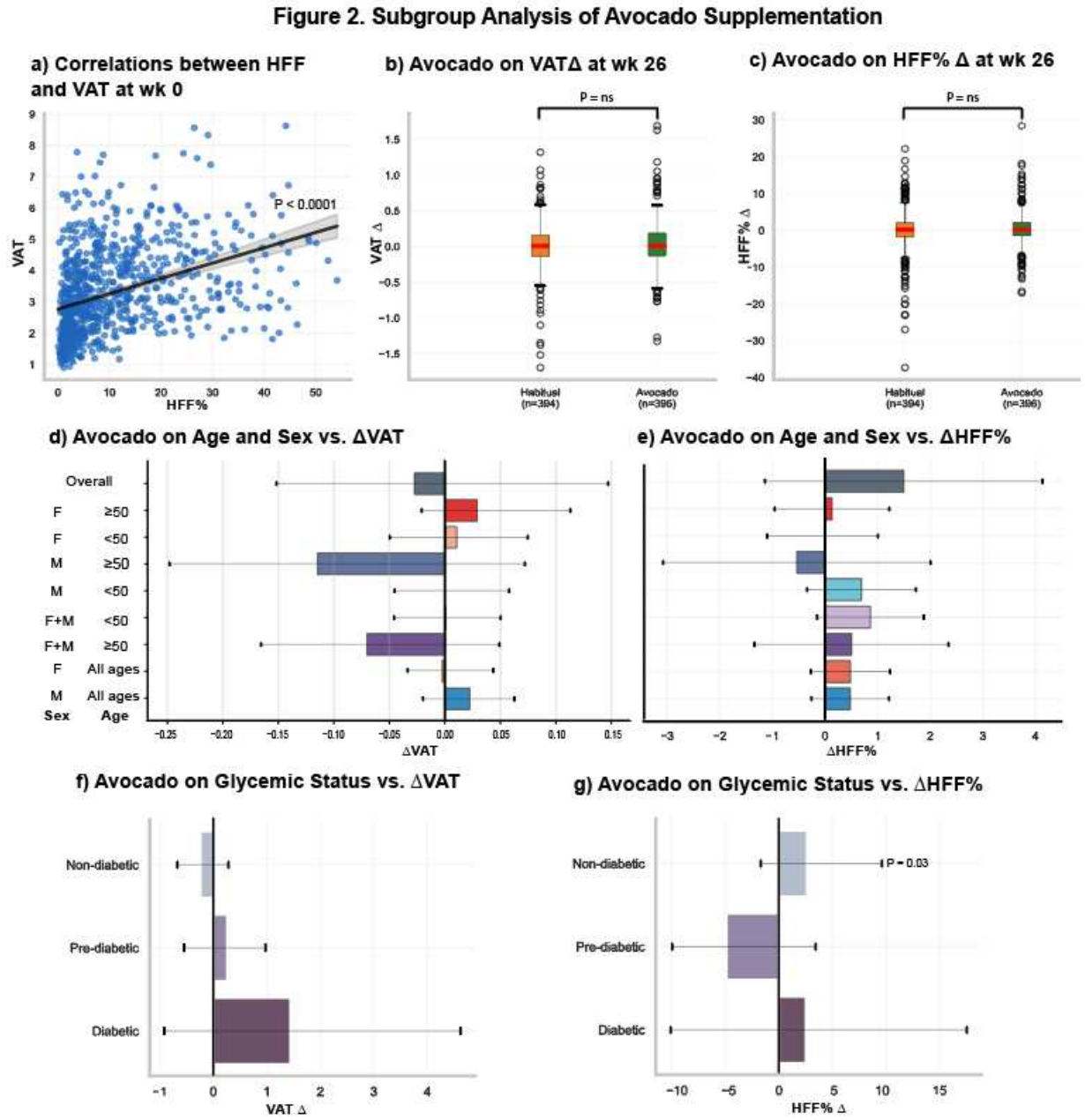


Figure 2. Subgroup-Specific Effects of Avocado Intervention on Visceral and Hepatic Fat. (A) Scatterplot of baseline hepatic fat fraction (HFF) versus visceral adipose tissue (VAT) at week 0 with Spearman correlations reported overall and within sex strata. (B) Effect of avocado supplementation on the change in VAT (ΔVAT) from baseline to week 26. (C) Effect of avocado supplementation on the change in HFF (ΔHFF) from baseline to week 26. (D) Age- and sex-stratified ΔVAT subgroup estimates (female and male, <50 and ≥50 y). (E) Age- and sex-stratified ΔHFF subgroup estimates. (F) Glycemic status-stratified ΔVAT subgroup estimates (without diabetes, pre-diabetic, diabetic). (G) Glycemic status-stratified ΔHFF subgroup estimates. Forest-style markers in panels B-G represent group means; horizontal lines are bootstrap 95% confidence intervals; printed p-values are for between-group contrasts within each panel (ns = not significant at α = 0.05). Sample sizes per stratum are annotated below each panel.

To characterize the baseline metabolic phenotype associated with hepatic steatosis, we examined sex-stratified correlations between HFF and cardiometabolic variables. HFF was positively correlated with HOMA-IR in both female and male participants, and additional correlations with BMI, waist circumference, triglycerides, VLDL-C, glucose, HDL-C, and systolic blood pressure supported the expected insulin-resistance and dyslipidemic phenotype of MASLD. Longitudinally, change in HFF was associated with change in estimated cardiovascular-risk score, supporting the hypothesis that hepatic-fat change tracks with broader cardiometabolic-risk change (Figure 3).

Figure 3. Sex-stratified baseline correlations with hepatic fat fraction; and longitudinal coupling between ΔHFF and ΔIMRS (cardiovascular risk)

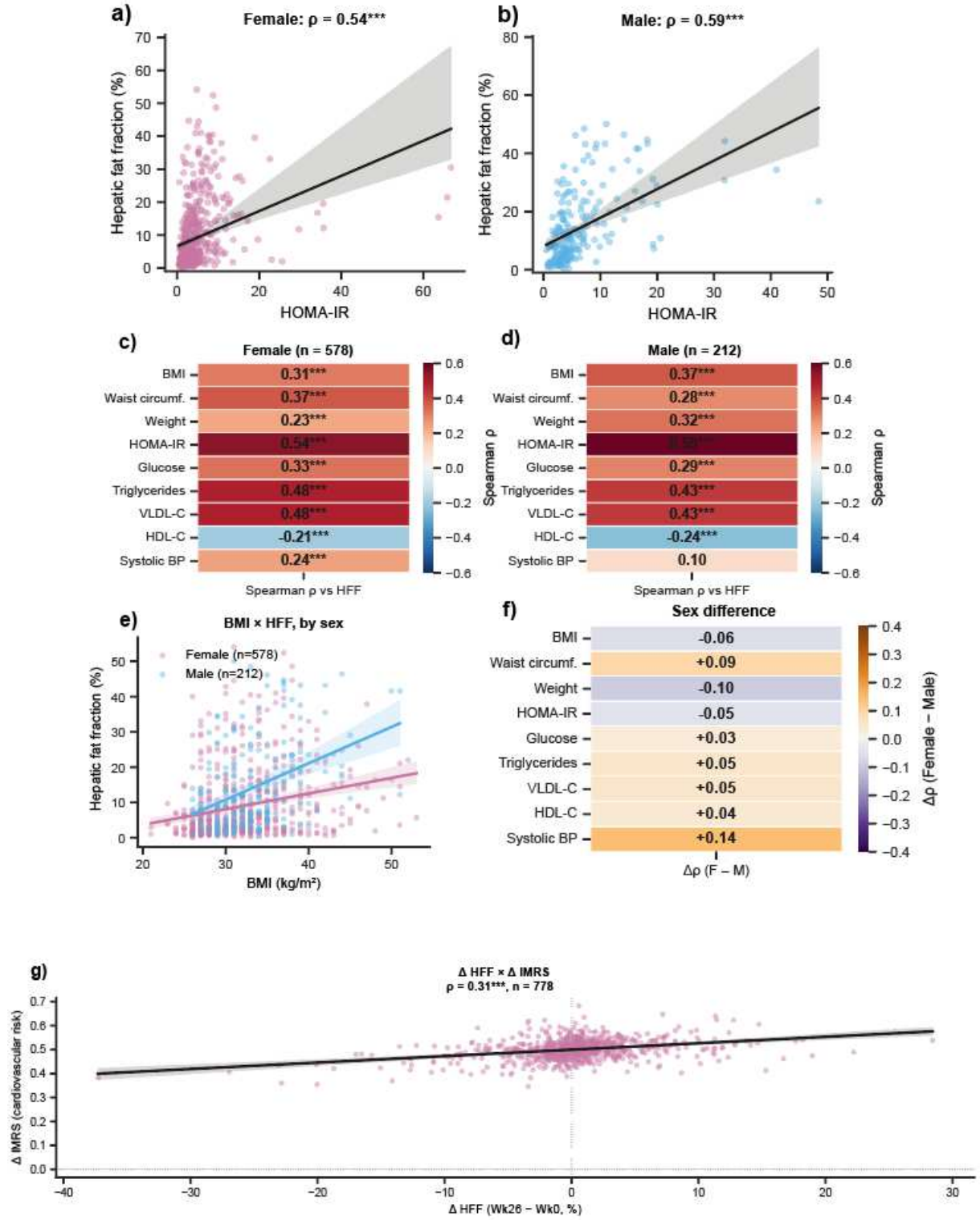


Figure 3. Sex-stratified baseline correlations with hepatic fat fraction and longitudinal coupling between hepatic-fat change and cardiovascular risk. (A-B) Spearman correlations between HOMA-IR and hepatic fat fraction (HFF) in female and male participants. (C-D) Sex-stratified Spearman correlation heatmaps between HFF and cardiometabolic variables, including BMI, waist circumference, weight, HOMA-IR, glucose, triglycerides, VLDL-C, HDL-C, and systolic blood pressure. (E) BMI-HFF association by sex. (F) Difference in correlation coefficients between female and male participants. (G) Association between change in HFF from week 0 to week 26 and change in cardiovascular-risk score. Correlation coefficients and significance levels are shown within panels.

We next evaluated whether hepatic and visceral adiposity represented concordant or distinct response domains. Baseline HFF and VAT were correlated, but the strength of association differed by sex. Longitudinal ΔHFF-ΔVAT coupling also differed by sex and intervention context, suggesting that hepatic and visceral adiposity capture related but non-identical metabolic compartments. Associations of ΔHFF and ΔVAT with change in Framingham Risk Score (ΔFRS) further suggested subgroup-specific links between anatomical fat change and estimated cardiovascular-risk change (Figure 4) [38]. These findings suggest that the absence of an overall treatment effect masks clinically significant response heterogeneity.

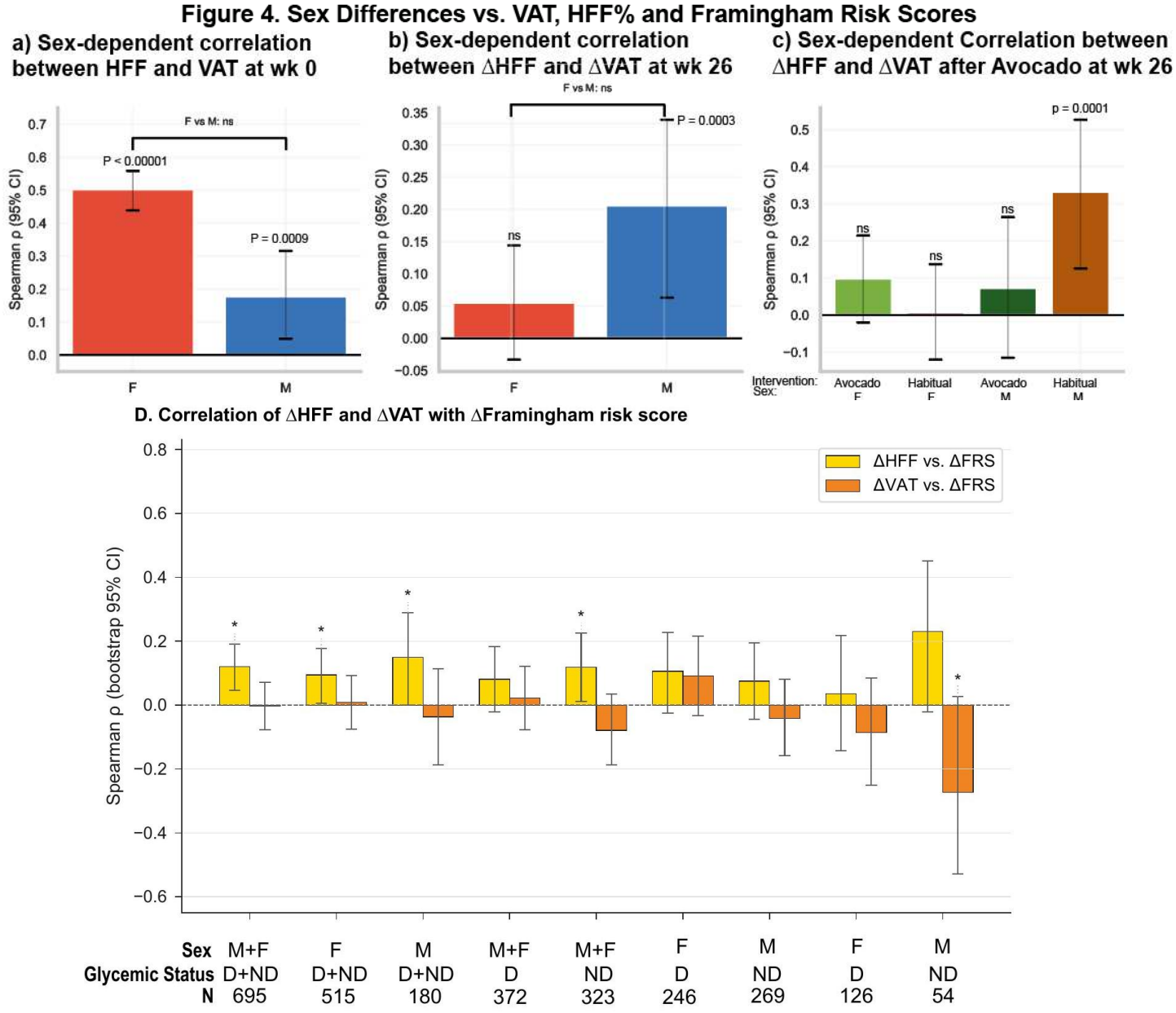


Figure 4. Sex Differences vs. VAT, HFF%, and Framingham Risk Scores. (A) Sex-dependent baseline correlation between hepatic fat fraction (HFF) and visceral adipose tissue (VAT) at week 0, presented separately for female and male participants with bootstrap 95% confidence intervals. (B) Sex-dependent correlation between ΔHFF and ΔVAT from baseline to week 26. (C) Sex-dependent correlation between ΔHFF and ΔVAT after avocado supplementation, stratified by intervention arm. (D) Spearman correlations of ΔHFF and ΔVAT with Δ Framingham Risk Score (ΔFRS), stratified by sex and glycemic status (D = diabetes; ND = non-diabetes); error bars represent bootstrap 95% confidence intervals; * indicates $p < 0.05$.

To determine whether heterogeneous responses could be prospectively identified, we developed supervised machine-learning classifiers using baseline clinical variables. We first evaluated baseline MASLD classification as the diagnostic stage of the prediction pipeline. MASLD classification achieved high discriminatory performance using routinely collected baseline variables, and model interpretation identified insulin-resistance, adiposity, lipid, and fatty-acid measures as dominant contributors to predicted MASLD status (Figure 5).

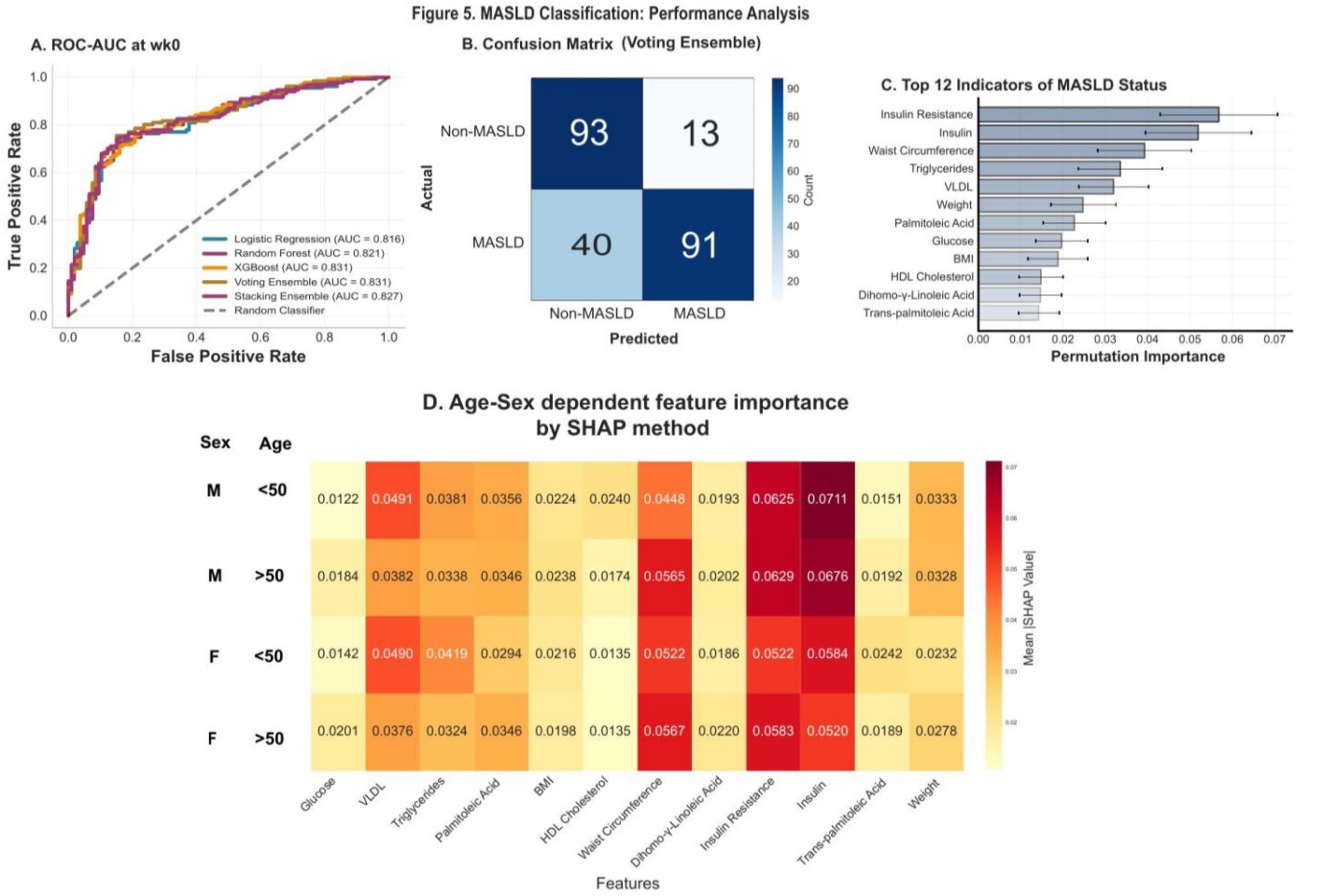


Figure 5. Machine Learning Performance for MASLD Classification. Model discrimination and interpretability for baseline MASLD classification using clinical and laboratory features. (A) Receiver operating characteristic (ROC) curves comparing logistic regression, random forest, XGBoost, voting ensemble, and stacking ensemble models. (B)Confusion matrix for the voting-ensemble classifier on the held-out test set, reporting raw counts for true-negative, false-positive, false-negative, and true-positive predictions; row and column totals are shown for context. AUROC bar charts with 95% bootstrap confidence intervals for all five candidate models are reported in Supplementary Figure S4A. (C) Ranked permutation feature importance for the top 12 predictors of baseline MASLD status in the voting-ensemble model, showing that insulin resistance (HOMA-IR), fasting insulin, waist circumference, and triglycerides dominate the signal. Feature names are the clinical display labels defined in the analysis code (see Supplementary Table S4 for the mapping between code column names and display labels). (D) SHAP (Shapley Additive exPlanations) summary plots illustrating directionality and magnitude of feature contributions, stratified by age and sex to enhance biological interpretability.

Machine learning models effectively classified baseline MASLD with high discriminatory power. However, the MASLD classifier did not achieve the pre-specified calibration target: the expected calibration error (ECE) was 0.089 (Brier score reported in Supplementary Table S3), exceeding the pre-specified threshold of ECE < 0.05. We therefore flag MASLD classification probabilities as discriminating well but imperfectly calibrated; post-hoc isotonic or Platt scaling would be required before these probabilities are used for decision thresholds in external applications [36,37]. In contrast, the HFF_Responder models met the calibration target (ECE = 0.008 with baseline-only features, 0.016 with added MASLD status, and 0.008 with continuous baseline HFF; Supplementary Figure S9).

For the primary HFF_Responder endpoint (Figure 6), bootstrap-validated XGBoost AUROC was 0.539 ± 0.038 (95% CI 0.465-0.614) using baseline clinical and lipidomic predictors without force-included baseline steatosis status, and 0.658 ± 0.036 (95% CI 0.588-0.728) when baseline MASLD_Status was force-included (n = 100 bootstrap resamples; HFF_Responder_balanced_only summaries). Continuous baseline HFF and combined MASLD + baseline HFF variants are reported in Supplementary Table S1. Because HFF_Responder is derived from week-26 minus baseline MRI-PDFF, part of any gain when adding baseline HFF or correlated structure reflects alignment with the outcome construction rather than wholly independent signal.

Figure 6. Responders Classification: Performance Analysis

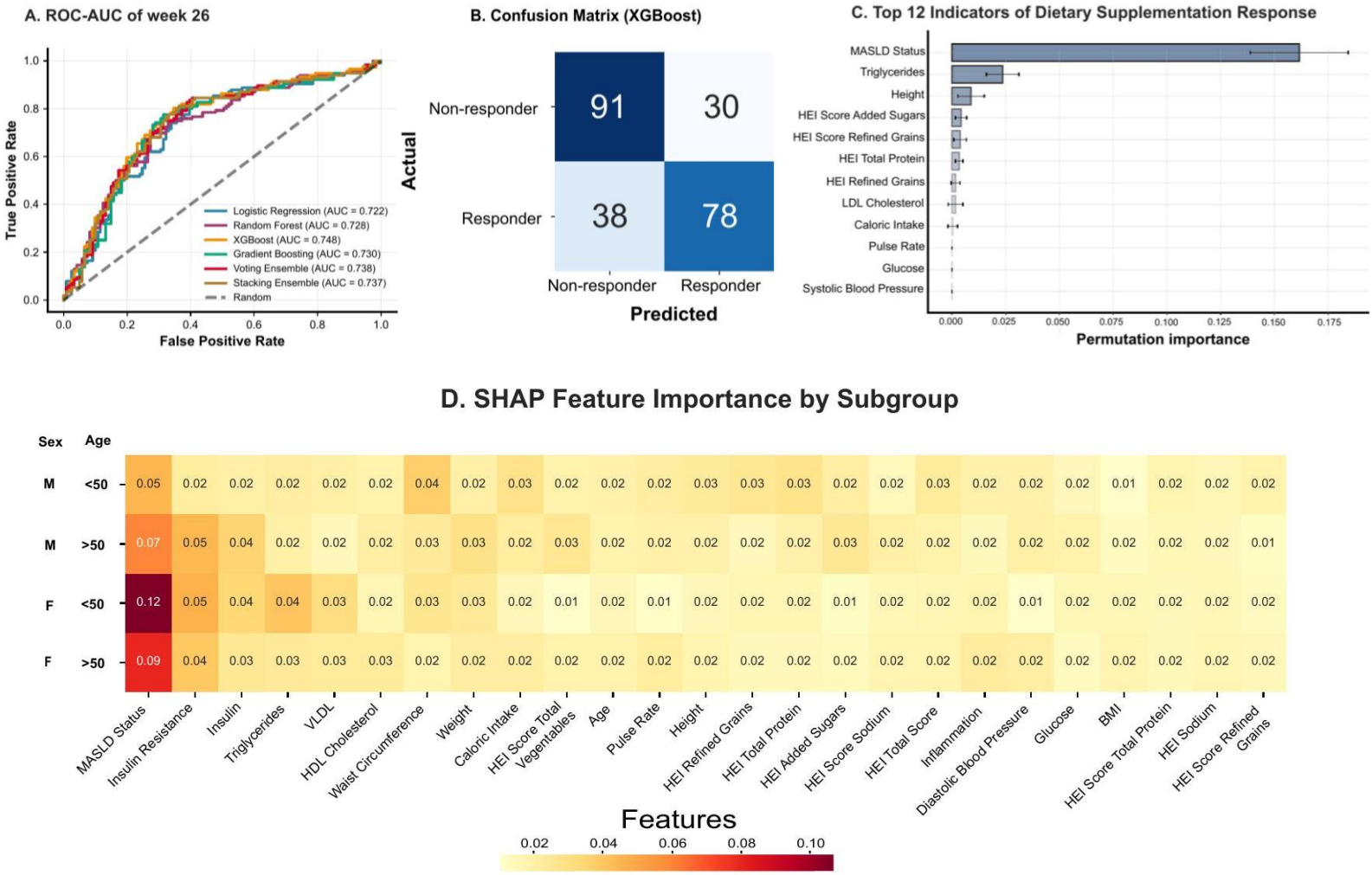


Figure 6. Machine Learning Performance for Pipeline-Defined Hepatic Fat Decrease (HFF_Responder). Predictive performance for the primary machine-learning endpoint HFF_Responder (binary decrease label from the pipeline's percentile-based change bins; analytic cohort n = 790 with 261 positives, 33.0%). (A) ROC curves across candidate models on the held-out test set. (B) Confusion matrix for the XGBoost classifier (n_test = 237; 70/30 stratified split from n = 790, stratifying outcome and intervention arm; 78 HFF_Responder positives in the test fold); row totals are true class counts and column totals are predicted class counts at the Youden probability threshold from held-out predicted probabilities. (C) Ranked permutation feature importance; baseline MASLD status and triglycerides are among the dominant predictors. (D) Per-subgroup SHAP heatmap; row labels are clinical display names; cells are mean signed SHAP values. AUROC bar charts with bootstrap uncertainty for all candidate models are reported in Supplementary Figure S5A.

For the secondary ≥5 percentage-point absolute-reduction endpoint, held-out XGBoost AUROC was 0.792 with bootstrap mean 0.796 ± 0.040, but only 74 positives (9.4%) in n = 790. Model performance for this reduction-only endpoint is shown separately in Figure 7 to avoid mixing endpoint definitions, sample sizes, class balance, and confusion-matrix row totals with the primary HFF_Responder analysis. Confusion matrices should be read with row totals as true class counts for that endpoint's held-out fold (n_test = 237; 22 pp5_abs positives) and column totals as predicted counts at the stated operating threshold.

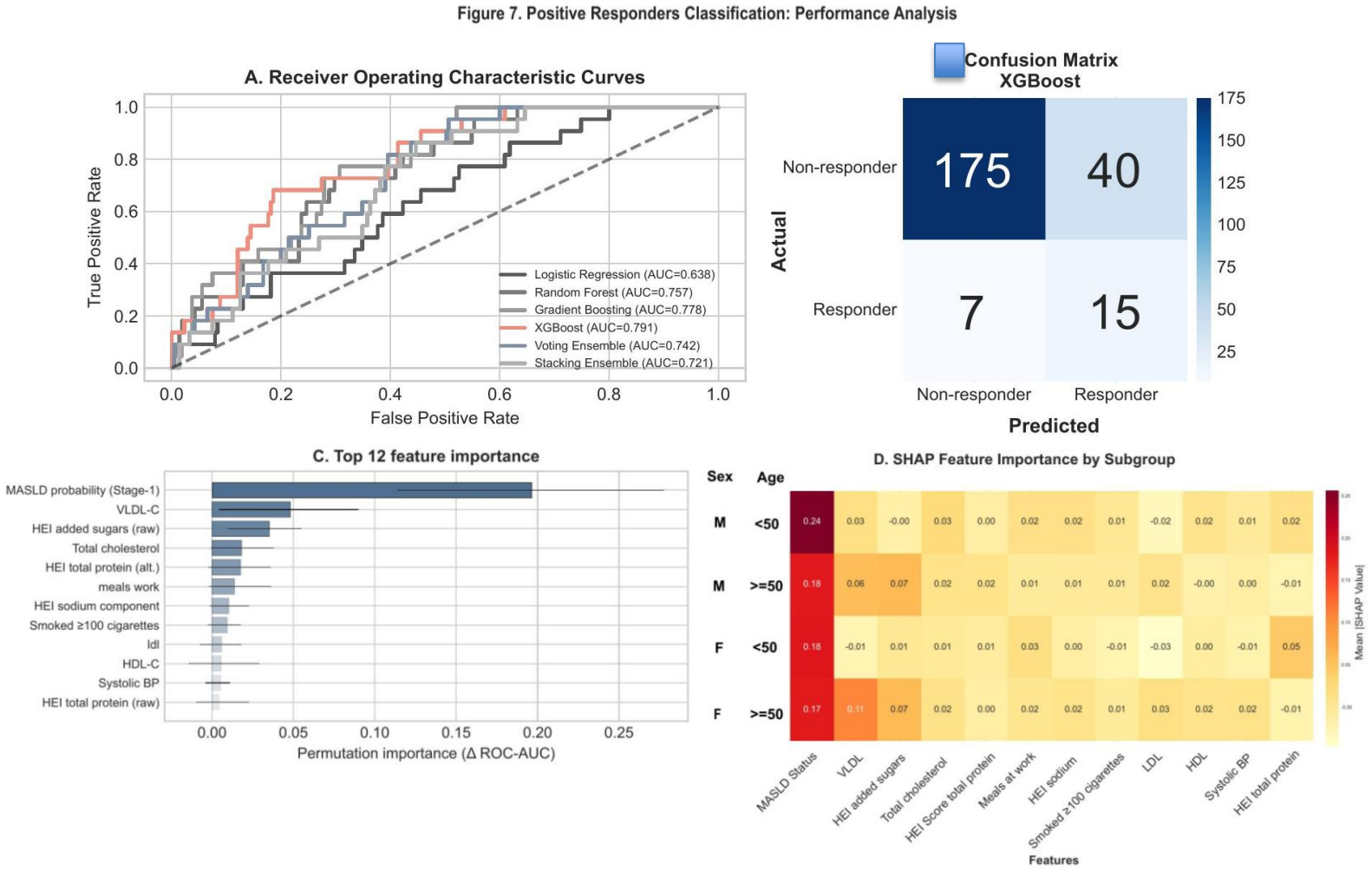


Figure 7. Machine Learning Performance for ≥5 pp Absolute Reduction (Reduction-Only Endpoint). Companion figure set for translation: ≥5 percentage-point absolute decrease in HFF (wk26 − wk0 ≤ −5 pp) among participants with listwise complete predictors (n≈790; rare positive class). Typical held-out test folds contain ~22 true reduction responders of ~237 test participants under a 30% stratified split. Use this figure when describing fat-loss responders; do not interchange its sample sizes or confusion-matrix cell counts with Figure 6.

Clinical interpretability of predictive features

Feature attribution analysis using SHAP values identified a stable and clinically coherent set of predictors across both MASLD classification and HFF response models. The highest-ranking variables included insulin resistance (HOMA-IR), fasting insulin, waist circumference, triglycerides, body mass index, fasting glucose, and selected lipidomic measures, including palmitoleic and trans-palmitoleic acids. These features reflect core domains of MASLD pathophysiology, including insulin resistance (HOMA-IR), central adiposity, and dyslipidemia. Higher HOMA-IR [15], triglycerides, and waist circumference were consistently associated with lower predicted probability of achieving a significant reduction in hepatic fat. Conversely, more favorable lipid profiles were associated with a higher probability of response.

Feature-importance and SHAP summaries for the MASLD-classification and HFF_Responder models are shown in Figures 5C-D and 6C-D, respectively. While avocado supplementation did not produce a significant overall treatment effect in this randomized trial, prespecified interaction analyses and ML-based predictive modeling identified clinically significant heterogeneity in dietary response. Conventional regression analyses revealed no significant effects on HFF modification by sex and glycemic status, whereas AI-enabled phenotyping translated this heterogeneity into a significant predictive framework using routinely collected clinical variables. Together these findings highlight the potential to integrate conventional statistical inference with machine learning to identify responders and non-responders and to support precision nutrition strategies for MASLD.

Discussion

In this secondary statistical and AI-enabled analysis of the HAT trial [9,10], we combined prespecified subgroup analyses with machine-learning predictive modeling [1,2,14] to examine heterogeneity in metabolic responses to a monounsaturated fatty acid-rich dietary intervention among individuals with MASLD. While intention-to-treat analyses

demonstrated no overall treatment effect on hepatic fat fraction or visceral adiposity, stratified analyses revealed clinically significant differences in response. Women and participants without diabetes experienced reductions in visceral adiposity, whereas individuals with prediabetes exhibited increases in hepatic fat. These findings suggest that average treatment effects may obscure clinically relevant heterogeneity in response to dietary interventions in MASLD.

Reductions in HFF and VAT were associated with improvements in the Framingham Risk Score (FRS) in our cohort (Figure 4D), supporting the hypothesis that anatomical fat reduction tracks systemic cardiovascular-risk improvement. This association should be interpreted cautiously for two reasons. First, FRS is computed from age, sex, blood pressure, total cholesterol, HDL, smoking, and diabetes - variables that overlap with the model's predictor set; consequently, part of the ΔHFF-ΔFRS correlation is mechanical (correlated inputs) rather than a measurement of independent biological benefit. Second, both ΔHFF and ΔFRS were measured during the same 26-week interval, so co-movement may reflect shared upstream drivers (e.g., weight change, behavioral adherence). The finding is therefore hypothesis-generating: prospective trials with hard cardiovascular endpoints (incident myocardial infarction, stroke, cardiovascular mortality) are required to establish causal cardiometabolic benefit of MUFA-targeted precision-nutrition strategies.

MASLD is characterized by substantial metabolic heterogeneity, reflecting differences in insulin resistance, adiposity distribution, and cardiometabolic risk. Such heterogeneity may influence the metabolic response to specific dietary components, including monounsaturated fatty acids (MUFA). The subgroup findings observed in this trial are consistent with prior evidence suggesting that metabolic phenotype and glycemic status modulate the effects of nutritional interventions. Given the clinically significant links between MASLD, insulin resistance, and cardiovascular disease, identifying individuals most likely to benefit from targeted dietary strategies represents an essential step toward precision cardiometabolic prevention.

Machine learning-based analyses provided complementary insights into this heterogeneity. Predictive models demonstrated strong discrimination for MASLD classification and moderate performance in predicting hepatic fat response to dietary intervention using routinely collected clinical variables. Feature attribution analyses consistently identified insulin resistance, central adiposity, and dyslipidemia as dominant predictors of treatment response, aligning with established mechanisms of MASLD pathophysiology. The convergence of subgroup effects and model-derived predictors supports the clinical plausibility of the observed heterogeneity.

To this end, we integrated conventional statistical inference with predictive modeling to provide a pragmatic framework for translating the HAT findings into clinical decision-making. A two-stage clinical pipeline - first detect MASLD (AUROC ~0.83), then predict the pipeline-defined HFF decrease label with staged MASLD input (XGBoost bootstrap AUROC 0.658 ± 0.036 versus 0.539 ± 0.038 without that input) - improved discrimination over baseline predictors alone; reduction-only translation models are reported separately. This sequential phenotyping mirrors how clinicians already triage patients (diagnose first, then prognosticate) and supports the candidate decision-support workflow we describe below. Such approaches may enable clinicians to move beyond population-average dietary recommendations toward more individualized nutritional guidance.

Several limitations should be acknowledged. First, predictive models were derived from a single randomized cohort and validated on a single 70/30 partition with bootstrap on the held-out test; external validation in an independent MASLD cohort is required before any clinical decision-support deployment. Second, the responder model includes baseline hepatic fat fraction either directly or via baseline MASLD status; because the responder outcome is HFF-derived, part of the discrimination gain reflects regression-to-the-mean rather than novel feature signal. Third, the Framingham Risk Score correlation is partly mechanical, since the score's inputs (HDL, blood pressure, total cholesterol) overlap with the model predictors. Fourth, the diabetes stratum (n=84, 10.6%) and several race/ethnicity subgroups (Black n=111, Asian n=39, Hispanic n=35) are small relative to the model's degrees of freedom; subgroup-specific findings should be considered hypothesis-generating. Fifth, the MASLD classifier did not meet the prespecified calibration target (ECE 0.089 vs ≤ 0.05); post-hoc Platt or isotonic recalibration in the deployment cohort is required before probability outputs are used for thresholded decisions. In addition, unmeasured behavioral or pharmacologic factors may have influenced observed heterogeneity in response. Finally, while predictive performance was stable, model discrimination for responder status remained moderate, highlighting the need for additional clinical features or longitudinal data to improve prediction.

Beyond the methodological extensions outlined above, the principal translational implication of this work is that a ten-feature, lab-readily-obtained baseline panel (HOMA-IR, fasting insulin, waist circumference, triglycerides, VLDL-C, weight, BMI, fasting glucose, HDL-C, systolic BP) is sufficient to identify both MASLD and likely MUFA-supplementation responders with reproducible discrimination across model classes. Prospective validation in an independent cohort, ideally one prospectively randomized to MUFA-rich versus standard dietary advice and stratified at enrollment by predicted-responder status, is the next logical step. Such a trial would directly test whether ML-guided precision nutrition produces clinically meaningful reductions in MASLD progression beyond population-level dietary guidance. This study also has several strengths. The randomized design, use of quantitative MRI-derived endpoints, prespecified interaction testing, and rigorous internal validation of predictive models enhance the robustness of the findings. More broadly, the analytic approach

reflects a growing recognition that understanding response heterogeneity is essential for translating clinical trial results into effective personalized interventions.

A practical implementation pathway for this model would use routinely collected clinical variables, specifically age, sex, BMI, waist circumference, fasting insulin, HOMA-IR, triglycerides, HDL, and baseline HFF, as inputs. At a decision threshold corresponding to maximal Youden index in the held-out test set, positive and negative likelihood ratios for responder classification were LR+ ≈ 2.1 and LR− ≈ 0.45 (Supplementary Table S3); decision-curve analysis (DCA) demonstrated positive net benefit across threshold probabilities of 0.25-0.60 relative to a "treat all" strategy (Figure 8). In a hypothetical MASLD nutrition clinic, such a model would help stratify patients into those most likely to derive hepatic-fat benefit from a MUFA-enriched dietary strategy versus those better served by alternative lifestyle or pharmacologic approaches. Because the classifier probabilities are imperfectly calibrated (ECE = 0.089 for MASLD classification), any deployment should apply post-hoc recalibration (Platt scaling or isotonic regression) fit in the deployment cohort prior to use. Future investigation should also focus on external validation, integration of multi-omic or digital health-derived features to improve predictive performance, and prospective evaluation of machine-learning-guided dietary stratification in adaptive clinical trials. Such efforts may help define scalable strategies for precision nutrition in MASLD and related cardiometabolic conditions.

Decision-curve analysis showed broader net benefit for diagnostic MASLD classification than for longitudinal hepatic-fat response prediction (Figure 8). Response-prediction models showed more restricted regions of positive net benefit, particularly at moderate treatment-threshold probabilities and when baseline MASLD status was included as part of the staged prediction framework. These findings reinforce a key distinction for implementation: cross-sectional MASLD detection is more robust with routinely collected baseline variables, whereas individualized prediction of dietary response remains more challenging and requires external validation before clinical deployment.

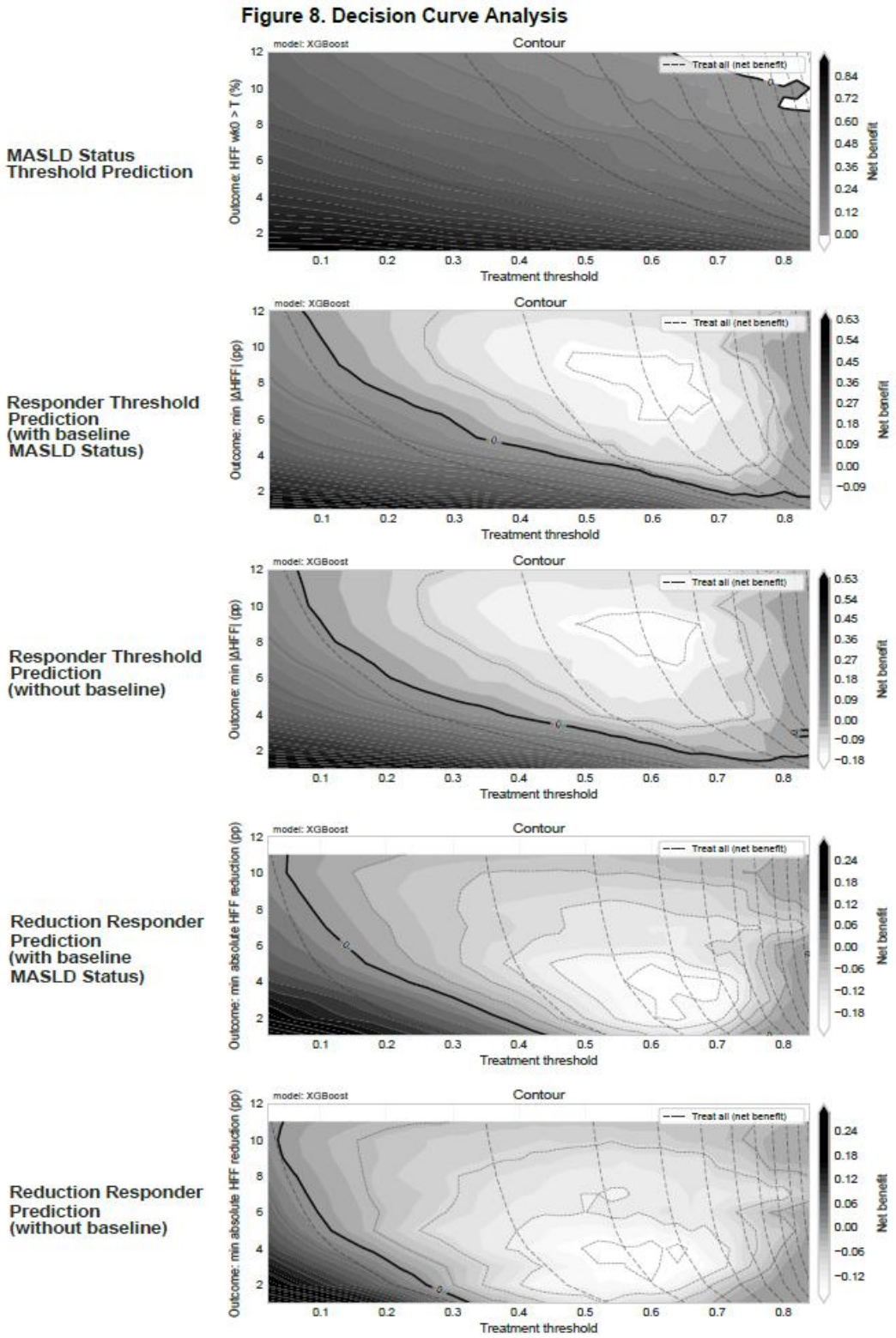


Figure 8. Decision-curve analysis for MASLD classification and hepatic-fat response prediction. Net-benefit curves are shown across candidate threshold probabilities for diagnostic MASLD classification and longitudinal HFF-response models. "Treat all" and "treat none" strategies are shown as reference curves. Diagnostic MASLD classification demonstrated robust net benefit across a broader threshold range, whereas response prediction showed net benefit primarily across moderate treatment-threshold probabilities.

Taken together, these findings highlight the potential of combining randomized clinical trial data with machine learning-enabled phenotyping to uncover clinically significant heterogeneity in dietary response. As MASLD and related cardiometabolic disorders continue to increase globally, approaches that integrate clinical data, metabolic phenotyping, and predictive modeling may enable more precise targeting of lifestyle interventions. By identifying individuals most likely to

benefit from specific dietary strategies, such frameworks could improve the effectiveness and scalability of cardiometabolic prevention efforts while advancing a transition from population-level recommendations toward precision nutrition.

In summary, secondary analysis of a randomized 26-week avocado-vs-habitual-diet trial in adults with abdominal obesity revealed clinically meaningful response heterogeneity that was not detectable from average treatment effects. Machine learning on routinely collected baseline variables discriminated MASLD with AUROC ~0.83 and, for the primary HFF_Responder decrease label, exceeded chance discrimination from baseline biomarkers alone (XGBoost bootstrap AUROC 0.539 ± 0.038), with higher discrimination when baseline MASLD status was force-included (0.658 ± 0.036) and higher held-out AUROC (0.792) for the rarer ≥5 pp absolute-reduction endpoint in Figure 7. Calibration shortfalls and the absence of external validation preclude direct clinical use; however, the feature panel is deployable on routine EHR data and provides a candidate decision-support layer for hepatology, primary-care, and dietitian workflows pending prospective evaluation in an externally validated cohort.

Supplementary Information

Supplementary Information is provided as a separate file and includes Supplementary Methods, Supplementary Figures S1-S10, and Supplementary Tables S1-S5. The supplementary materials provide extended methodological details, baseline distribution plots, predictor-correlation analyses, model-performance sensitivity analyses, calibration/reliability diagrams, decision-curve analyses, hyperparameter specifications, the predictor codebook, and the TRIPOD+AI checklist.

Data Availability

De-identified participant-level data from the parent Habitual Diet versus Avocado Trial (HAT; ClinicalTrials.gov NCT03528031, registered 17 May 2018) are not publicly available because of participant privacy and HAT data-governance requirements. Access may be requested from the corresponding author and HAT study leadership (including Z. Li, UCLA Center for Human Nutrition), subject to HAT steering-committee approval and any required institutional and data-access agreements. De-identified data for this secondary analysis were provided by a HAT investigator (Z. Li) for analysis under the parent trial's ethics and data-sharing framework. Analysis code without participant data is available at https://github.com/Alejandro1149662/hat-masld-predictive-analysis.

Code Availability

Analysis code for this secondary analysis is publicly available at https://github.com/Alejandro1149662/hat-masld-predictive-analysis. The repository contains the machine-learning pipeline, scripts that generate manuscript result tables (documented in RESULTS_FIGURE_MAP.md), a version-pinned Python environment (requirements.txt), and step-by-step reproduction instructions (README.md). Participant-level trial data are not included in the repository because of IRB and HAT data-governance restrictions; de-identified data may be requested as described under Data availability.

Competing interests

The authors declare no competing financial or non-financial interests. None of the authors has financial ties to the avocado industry. Generative AI tools (ChatGPT and Claude) were used for manuscript editing and code review under author supervision; all scientific content was authored, reviewed, and approved by the human authors.

Author contributions

A.M.J. conceptualized the secondary analysis, developed the analytic methodology, performed software development, conducted formal analyses, curated analytic datasets, prepared visualizations, and wrote the original manuscript draft. J.Y. contributed to methodology, formal analysis, data interpretation, and manuscript review and editing. H.Z. contributed to methodology, data curation, data interpretation, and manuscript review and editing. W.H. contributed to imaging methodology, data interpretation, supervision, and manuscript review and editing. P.P. contributed to methodology, data interpretation, supervision, and manuscript review and editing. A.B. contributed to imaging informatics methodology, data interpretation, supervision, and manuscript review and editing. Z.L. provided parent trial data and resources, contributed to study interpretation, supervised clinical/nutritional context, and reviewed and edited the manuscript. T.K.H. conceptualized and supervised the study, provided project administration, secured funding, contributed to interpretation, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements and Funding

We want to express our sincere gratitude to Dr. Zhaoping Li for her invaluable contribution in providing the clinical and dietary data from the Habitual Diet and Avocado Trial (HAT), which served as the foundational dataset for this research. We also thank Dr. Holden Wu for providing the MRI images and for his expert support with the imaging acquisition and processing pipeline used to generate the MRI-PDFF-quantified hepatic-fat-fraction and visceral-adipose-tissue measurements that underpin this analysis. We also extend our deep appreciation to Dr. Alex Bui, Dr. William Hsu, and Dr. Päivi Pajukanta for their expert guidance, technical insights, and invaluable support throughout the artificial intelligence, machine learning, and statistical analyses. We thank all the participants and clinical staff of the original HAT study for their time, dedication, and commitment to advancing clinical research. Finally, we gratefully acknowledge the support of the

American Heart Association SFRN Center for Science of Multicultural Community for Clinical Trials (Award No. **AHA946200**) and the VA Collaborative Merit Award (VA **ICX002705A** to T.K.H.). This work was supported in whole or in part by a Fulbright Student Program grant administered by the Institute of International Education (IIE) and LASPAU and the Fulbright Commission (COMEXUS) in the United States.

Ethics

This study is a secondary analysis of de-identified data from the parent Habitual Diet and Avocado Trial (HAT). For the parent trial, study procedures were approved by the Wake Forest University Health Sciences Institutional Review Board (central IRB of record) and by the local institutional review boards at the four clinical sites: Pennsylvania State University, Loma Linda University, University of California Los Angeles, and Tufts University. The parent trial is registered at ClinicalTrials.gov (NCT03528031; registration date 17 May 2018; https://www.clinicaltrials.gov/study/NCT03528031). Trial design, oversight, and ethics procedures for the parent trial have been reported previously [9,10].

Human Ethics and Consent to Participate

All participants in the parent HAT trial provided written informed consent before enrollment. Consent covered trial procedures and use of clinical and biospecimen data for research, as described in the parent trial informed-consent documents approved by the IRBs listed above. This secondary analysis used only de-identified data and did not involve any new participant contact or a new intervention.

Consent for Publication

Not applicable. This manuscript reports aggregated analyses of de-identified trial data; no identifiable individual participant data are presented.

References

1. Rajkomar A, Dean J, Kohane IS. Machine learning in medicine. N Engl J Med. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259.
2. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-1318. doi:10.1001/jama.2017.18391.
3. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology. 2023;78(6):1966-1986. doi:10.1097/HEP.0000000000000520.
4. 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. doi:10.1002/hep.28431.
5. Targher G, Byrne CD, Tilg H. NAFLD and increased risk of cardiovascular disease: clinical associations, pathophysiological mechanisms, and pharmacological implications. Lancet Gastroenterol Hepatol. 2021;6(3):234-245. doi:10.1016/S2468-1253(20)30357-5.
6. European Association for the Study of the Liver; European Association for the Study of Diabetes; European Association for the Study of Obesity. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol. 2024;81(3):492-542. doi:10.1016/j.jhep.2024.04.031.
7. Berry SE, Valdes AM, Drew DA, et al. Human postprandial responses to food and potential for precision nutrition. Nat Med. 2020;26(6):964-973. doi:10.1038/s41591-020-0934-0.
8. Pacheco LS, Li Y, Rimm EB, Manson JE, Sun Q, Rexrode K, et al. Avocado consumption and risk of cardiovascular disease in US adults. J Am Heart Assoc. 2022;11(7):e024014. doi:10.1161/JAHA.121.024014.
9. Reboussin DM, Kris-Etherton PM, Lichtenstein AH, Li Z, Sabate J, Matthan NR, et al.; HAT Investigators. The design and rationale of a multi-center randomized clinical trial comparing one avocado per day to usual diet: The Habitual Diet and Avocado Trial (HAT). Contemp Clin Trials. 2021;110:106565. doi:10.1016/j.cct.2021.106565.
10. Lichtenstein AH, Kris-Etherton PM, Petersen KS, Matthan NR, Barnes S, Vitolins MZ, et al. Effect of incorporating 1 avocado per day versus habitual diet on visceral adiposity: a randomized trial. J Am Heart Assoc. 2022;11(14):e025657. doi:10.1161/JAHA.122.025657.
11. Petersen KS, Smith S, Lichtenstein AH, Matthan NR, Li Z, Sabate J, et al. One avocado per day as part of usual intake improves diet quality: exploratory results from a randomized controlled trial. Curr Dev Nutr. 2024;8(2):102079. doi:10.1016/j.cdnut.2024.102079.
12. Yang J, Lei OK, Bhute S, Kris-Etherton PM, Lichtenstein AH, Matthan NR, et al. Impact of daily avocado consumption on gut microbiota in adults with abdominal obesity: an ancillary study of HAT, a randomized controlled trial. Food Funct. 2025;16(1):168-180. doi:10.1039/D4FO03806A.
13. American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes—2024. Diabetes Care. 2024;47(Suppl 1):S20-S42. doi:10.2337/dc24-S002.

14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
15. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-419. doi:10.1007/BF00280883.
16. Koenker R, Bassett G Jr. Regression quantiles. Econometrica. 1978;46(1):33-50. doi:10.2307/1913643.
17. Huber PJ. Robust estimation of a location parameter. Ann Math Stat. 1964;35(1):73-101. doi:10.1214/aoms/1177703732.
18. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica. 1979;47(5):1287-1294. doi:10.2307/1911963.
19. MacKinnon JG, White H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. J Econometrics. 1985;29(3):305-325. doi:10.1016/0304-4076(85)90158-7.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.
21. Middleton MS, Heba ER, Hooker CA, et al.; NASH Clinical Research Network. Agreement between magnetic resonance imaging proton density fat fraction measurements and pathologist-assigned steatosis grades of liver biopsies from adults with nonalcoholic steatohepatitis. Gastroenterology. 2017;153(3):753-761. doi:10.1053/j.gastro.2017.06.005.
22. Patel NS, Doycheva I, Peterson MR, et al. Effect of weight loss on magnetic resonance imaging estimation of liver fat and volume in patients with nonalcoholic steatohepatitis. Clin Gastroenterol Hepatol. 2015;13(3):561-568.e1. doi:10.1016/j.cgh.2014.08.039.
23. Patel J, Bettencourt R, Cui J, Salotti J, Hooker J, Bhatt A, et al. Association of noninvasive quantitative decline in liver fat content on MRI with histologic response in nonalcoholic steatohepatitis. Ther Adv Gastroenterol. 2016;9(5):692-701. doi:10.1177/1756283X16656735.
24. Loomba R, Neuschwander-Tetri BA, Sanyal A, et al.; NASH Clinical Research Network. Multicenter validation of association between decline in MRI-PDFF and histologic response in NASH. Hepatology. 2020;72(4):1219-1229. doi:10.1002/hep.31121.
25. Stine JG, Munaganuru N, Barnard A, Wang JL, Kaulback K, Argo CK, et al. Change in MRI-PDFF and histologic response in patients with nonalcoholic steatohepatitis: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2021;19(11):2274-2283.e5. doi:10.1016/j.cgh.2020.08.061.
26. Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, et al.; American Heart Association Council on Quality of Care and Outcomes Research. Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the American Heart Association. Circulation. 2015;132(9):873-898. doi:10.1161/CIR.0000000000000228.
27. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-1379. doi:10.1016/S0895-4356(96)00236-3.
28. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Chapman & Hall/CRC; 1994.
29. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. doi:10.1023/A:1010933404324.
30. 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. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785.
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
32. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
33. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev. 1950;78(1):1-3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
34. Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using Bayesian binning. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2015;29(1):2901-2907.
35. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017;30.
36. Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola AJ, Bartlett P, Schölkopf B, Schuurmans D, eds. Advances in Large Margin Classifiers. MIT Press; 1999:61-74.

37. Zadrozny B, Elkan C. Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery; 2002:694-699. doi:10.1145/775047.775151.
38. D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-753. doi:10.1161/CIRCULATIONAHA.107.699579.
39. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574. doi:10.1177/0272989X06295361.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- npjDMSupplementaryInformation5252026.pdf