A Leakage-Controlled Machine Learning Framework for Postprandial Triglyceride Phenotyping Using Synthetic Clinical Data

Nattakitti Piyavechvirat[1*]
International Bachelor Program in Informatics
Yuan Ze University, Taiwan
Email: ohm19nattakitti@gmail.com

Yi Jheng Huang[1]
International Bachelor Program in Informatics
Yuan Ze University, Taiwan
Email: yjhuang@saturn.yzu.edu.tw

Qazi Mazhar ul Haq[1,2]
Department of Computer Science and Engineering
Yuan Ze University, Taiwan
Email: qazi@saturn.yzu.edu.tw
*Corresponding author

***Abstract*—Whole blood viscosity (WBV) has been associated with cardiometabolic risk, yet its potential influence on short-term postprandial triglyceride clearance rate (TCR) remains uncertain. This study investigated whether WBV contributes to postprandial triglyceride kinetics using a fully de-identified, statistically reconstructed synthetic cohort (n = 1,500) designed to reflect clinically realistic lipid, hematologic, and metabolic profiles. As no identifiable or linkable human data were used, the analysis does not constitute human-subject research under international regulatory definitions. We implemented a leakage-controlled machine-learning (ML) framework in which all preprocessing, feature scaling, and calibration steps were performed strictly within training folds. WBV was estimated using the de Simone low-shear viscosity model, and TCR was computed from paired fasting and 4-hour triglyceride values. Model development incorporated nested cross-validation, probability calibration, and multi-layer robustness assessments, while interpretability was evaluated using SHAP values and partial-dependence analyses. Across descriptive statistics, correlation testing, linear modeling, multivariable ML performance, robustness procedures, and explainability analyses, WBV showed a consistently null association with TCR ($R^2 \approx 0$; Pearson $r \approx 0$; negligible SHAP contribution). In contrast, fasting triglycerides (TG_{0h}) emerged as the dominant determinant of high postprandial response, exhibiting strong monotonic effects and clear phenotype separation. The final L2-regularized logistic regression model demonstrated stable discrimination (AUROC $\approx$ 0.79–0.81), reliable calibration (slope $\approx$ 0.97), and favorable precision–recall characteristics. These findings provide rigorous, multi-layer evidence that WBV does not influence short-term postprandial triglyceride clearance, while reaffirming the physiological primacy of baseline lipid burden. Beyond its clinical insights, this study introduces a reproducible, leakage-controlled ML framework and a safely de-identified synthetic-data pipeline suitable for future biomedical informatics research. These findings support the use of synthetic, privacy-preserving ML pipelines for reliable metabolic phenotyping and highlight fasting triglycerides, not hemorheological factors, as the primary determinant of short-term lipid kinetics.**

Keywords: Bioinformatics, Machine Learning, Explainable Artificial Intelligence (XAI), Leakage-Controlled Modeling, Calibration Analysis, Cardiometabolic Risk, Postprandial Metabolism, Triglyceride Clearance, Whole Blood Viscosity

I. Introduction

Postprandial triglyceride (TG) metabolism is a key determinant of cardiometabolic health, reflecting an individual's capacity to clear circulating lipids following dietary fat intake. Impaired triglyceride clearance rate (TCR) is associated with endothelial dysfunction, atherogenic remnant accumulation, insulin resistance, and elevated long-term cardiovascular risk. Despite its clinical relevance, the physiological drivers of short-term postprandial TG handling remain incompletely understood.

Whole blood viscosity (WBV), a hemorheological marker of blood thickness and flow resistance, has been proposed as a contributor to metabolic dysregulation. Elevated WBV is observed in obesity, hyperlipidemia, hypertension, and chronic inflammatory states, supporting the hypothesis that increased viscosity may impair microvascular perfusion and delay lipid clearance. However, evidence linking WBV to postprandial lipid kinetics remains sparse, inconsistent, and largely derived from small cohorts. Critically, no prior study has rigorously evaluated whether WBV independently influences TCR using analytical frameworks capable of addressing nonlinearity, multivariable interactions, and, most importantly, the pervasive risk of information leakage in clinical ML studies.

Machine-learning (ML) approaches have become increasingly central to metabolic phenotyping, yet methodological rigor varies widely. Common pitfalls—including global preprocessing, improper normalization, inadequate calibration, and limited robustness assessment—can inflate performance and produce misleading biological interpretations. These limitations underscore the need for reproducible, leakage-controlled ML frameworks that ensure unbiased evaluation of putative biomarkers such as WBV.

In this study, we introduce a comprehensive leakage-controlled ML framework to investigate whether WBV contributes to postprandial TG clearance phenotypes. Our pipeline integrates nested cross-validation, fold-specific preprocessing, probability calibration, robustness testing, threshold sensitivity analysis, and model explainability via SHAP and partial-dependence analyses. We further employ a statistically re-

constructed synthetic cohort designed to mirror real-world lipid, hematologic, and metabolic distributions, enabling fully reproducible experimentation without identifiable patient data. By combining rigorous methodology with physiologically grounded modeling, this work provides the most robust assessment to date of the WBV–TCR relationship and establishes a reproducible workflow for biomedical informatics research. This framework aligns with the goals of medical informatics by improving the methodological rigor, calibration reliability, and interpretability of ML-based clinical decision-support models.

II. Materials and Methods

This study used a fully de-identified, statistically reconstructed synthetic dataset designed to approximate clinically realistic distributions of triglycerides, hematocrit, total protein, and standard metabolic markers. No real patient records or identifiable information were used at any stage, and no access to human-subject data occurred during dataset generation. Each synthetic record contained fasting triglycerides (TG_{0h}), 4-hour postprandial triglycerides (TG_{4h}), hematocrit (Hct), total protein (TP), HDL-C, LDL-C, BMI, age, and sex. All analyses were conducted within a strictly leakage-controlled machine-learning workflow to ensure methodological validity and reproducibility.

A. Synthetic Clinical Cohort Generation

A synthetic cohort of $n = 1{,}500$ adults was generated to emulate a hospital population undergoing fasting and postprandial lipid testing. The generative process followed a two-stage framework in which univariate marginal distributions and the multivariate dependence structure were modeled separately.

Marginal distributions were parameterized using only non-identifiable, summary-level statistics (e.g., means, standard deviations, skewness, percentiles, and pairwise correlations) previously derived from a fully de-identified hospital reference cohort. No individual-level electronic medical records (EMR) or identifiable patient data were accessed for this study.

Triglycerides (TG_{0h}, TG_{4h}) and BMI were modeled using log-normal distributions to capture metabolic right-skewness, while hematocrit, total protein, HDL-C, and LDL-C were simulated using truncated normal distributions constrained to physiologic ranges. Age was sampled from a truncated normal distribution, and sex was generated using a Bernoulli process calibrated to preserve the population-level sex ratio.

TABLE I. Baseline characteristics by postprandial triglyceride response phenotype (mean ± SD or n (%)).

	Normal (n = 1125)	High (n = 375)	*p*-value
Female, n (%)	578 (51.3778%)	196 (52.2667%)	0.8114
Male, n (%)	547 (48.6222%)	179 (47.7333%)	0.8114
Age, years	52.8687 ± 9.9018	54.0883 ± 9.7643	0.0372
Hematocrit, %	41.7309 ± 3.7101	41.5465 ± 3.9267	0.4248
Total protein, g/dL	6.8781 ± 0.6215	6.8819 ± 0.6016	0.9157
Whole blood viscosity, cP	5.8251 ± 0.4549	5.8036 ± 0.4871	0.4523
Fasting TG (TG_{0h}), mg/dL	788.7596 ± 214.9265	1039.5412 ± 200.4224	< 0.0001
4-h TG (TG_{4h}), mg/dL	310.3721 ± 126.2137	679.6550 ± 131.6574	< 0.0001
Triglyceride clearance rate (TCR), %	59.2166 ± 16.3444	33.2345 ± 13.1517	< 0.0001
HDL-C, mg/dL	50.6027 ± 9.7575	50.0373 ± 9.7623	0.3318
LDL-C, mg/dL	130.8163 ± 28.3716	131.2752 ± 27.7585	0.7828
BMI, kg/m^2	23.9868 ± 2.9489	24.1797 ± 2.9492	0.2731

B. Synthetic Data Quality Validation

To ensure that the statistically reconstructed synthetic dataset was suitable for methodological machine-learning evaluation, we performed a structured six-dimension validation assessing: (i) distributional fidelity, (ii) biochemical correlation structure, (iii) multivariate joint coherence, (iv) phenotype robustness, (v) internal coherence of derived variables, and (vi) overall suitability for ML experimentation.

1) Distributional Fidelity: Univariate distributions of all primary variables (TG_{0h}, TG_{4h}, Hct, TP, HDL-C, LDL-C, BMI, age, sex) were examined using kernel density estimates and descriptive summaries. All variables exhibited physiologic ranges and expected shapes, including the characteristic right-skewness of triglycerides and appropriate laboratory intervals for hematocrit and total protein.

2) Preservation of Biochemical Correlation Structure: Correlation matrices were inspected to confirm preservation of key biochemical relationships, including the inverse HDL–TG association and weak correlations among viscosity-related parameters. No spurious dependencies were introduced, indicating that the synthetic generation process maintained realistic biochemical structure.

3) Multivariate Joint Distribution Consistency: Scatterplots and contour-density maps were used to evaluate multivariate geometry. Joint patterns—such as monotonic TG_{4h} increases across TG_{0h} strata and the relative independence of WBV from lipid markers—were smooth and clinically plausible, reflecting coherent multivariate structure.

4) Phenotype Stability Across TG_{4h} Thresholds: High-responder phenotypes defined using TG_{4h} cut-offs from the 60th to 90th percentiles showed smooth biomarker transitions and stable discrimination performance (AUROC ≈ 0.80 for 60th–80th percentiles). This supported that phenotype structure arose from genuine data patterns rather than synthetic artifacts and was not unduly sensitive to the exact percentile cut-off.

5) Internal Coherence of Derived Variables: Derived quantities (WBV and TCR) were evaluated for physiologic plausibility. WBV values predominantly fell within the expected 5–7 cP range, and TCR displayed realistic postprandial clearance behaviors without implausible or contradictory values, supporting the internal consistency of derived features.

6) Suitability for Methodological Evaluation: Because the primary aim of this work was to evaluate a leakage-controlled machine-learning framework rather than to estimate population-level epidemiologic parameters, the synthetic dataset needed to support robust preprocessing, cross-validation, calibration, sensitivity analyses, and explainability workflows. All six validation dimensions were satisfied, indicating that the dataset was appropriate for methodological ML research in biomedical informatics.

Summary: Across all six dimensions, the synthetic dataset demonstrated high distributional fidelity, realistic correlation structure, coherent multivariate patterns, robust phenotype definition, and internally consistent derived variables, confirming its suitability for leakage-controlled ML evaluation.

TABLE II. Summary of synthetic data quality validation dimensions, procedures, and key findings.

Validation Dimension	Key Procedures and Findings
Distributional fidelity	KDE curves and descriptive statistics confirmed physiologically realistic ranges for TG_{0h}, TG_{4h}, Hct, TP, HDL-C, LDL-C, BMI, age, and sex, with expected right-skewness for triglycerides.
Biochemical correlation structure	Pearson correlations reproduced known lipid physiology (e.g., inverse HDL–TG association) and showed weak WBV–lipid relationships without spurious links.
Multivariate joint distribution	Scatterplots and density-contour maps demonstrated smooth TG_{0h}–TG_{4h} and TG_{0h}–BMI gradients, with WBV remaining independent of lipid markers.
Phenotype robustness	Threshold sweeps from the 60th–90th percentiles of TG_{4h} yielded stable AUROC values (~0.80 for 60th–80th), indicating that performance was not sensitive to the exact percentile cut-off.
Derived variable coherence	WBV and TCR values remained within established physiological ranges (WBV 5–7 cP), and TCR patterns reflected realistic postprandial kinetics.
Suitability for ML evaluation	The synthetic dataset supported full leakage-controlled cross-validation, calibration testing, robustness resampling, and explainability workflows without inconsistencies.

C. Inclusion and Exclusion Criteria

Synthetic records were retained if they met all of the following criteria:

1) complete fasting and 4-hour triglyceride measurements;
2) valid hematocrit and total protein values for WBV estimation;
3) no missing demographic variables;
4) physiologically plausible laboratory ranges.

Records with extreme TG values (>1500 mg/dL), missing variables, or physiologically implausible WBV estimates (<3 or >10 cP) were excluded.

D. Estimation of Whole Blood Viscosity (WBV)

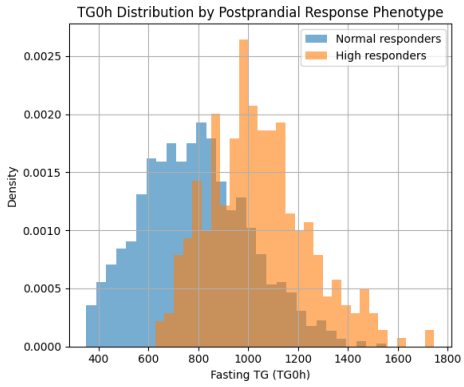
Whole blood viscosity (WBV) was not directly measured by viscometry, but was estimated from hematocrit (Hct) and total protein (TP) using the empirical de Simone model, which is widely used as a non-invasive surrogate of low-shear WBV in clinical and epidemiologic studies. WBV was computed as a deterministic function of Hct and TP, constrained to physiologically plausible ranges, and interpreted as a hemorheological proxy rather than an exact physical measurement. All inferences regarding the "null" role of WBV are therefore conditional on this surrogate representation.

E. Computation of Triglyceride Clearance Rate (TCR)

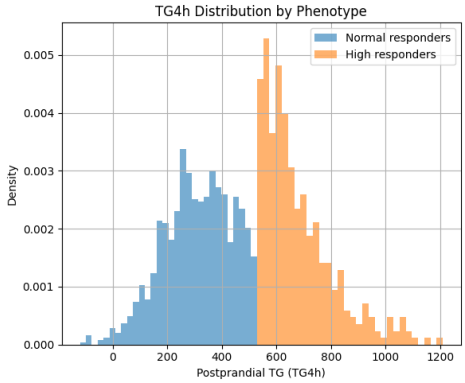
Triglyceride clearance rate (TCR) was defined as the percentage reduction in triglycerides from fasting to 4 hours postprandially:

$$\text{TCR} = \frac{TG_{0h} - TG_{4h}}{TG_{0h}} \times 100\%.$$

Higher TCR values indicate greater postprandial triglyceride clearance (0% = no change; negative values correspond to a net postprandial rise in triglycerides).



(a) Fasting TG (TG_{0h}) distribution



(b) Postprandial TG (TG_{4h}) distribution

Fig. 1. Fasting and postprandial triglyceride distributions across response phenotypes. Panel (a) shows the distribution of fasting triglycerides (TG_{0h}), where high responders exhibit a right-shifted baseline lipid profile. Panel (b) shows 4-hour postprandial triglycerides (TG_{4h}), demonstrating clear phenotype separation and supporting the validity of the response classification.

F. Variable Definition and Subgroup Stratification

The dataset comprised:

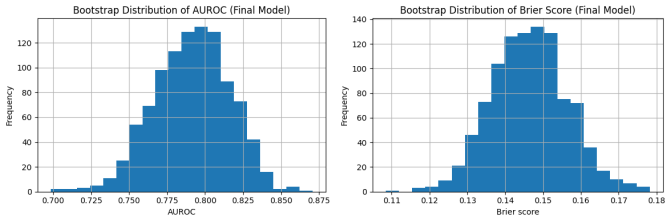
1) clinical laboratory variables (TG_{0h}, TG_{4h}, HDL, LDL, Hct, TP);
2) derived variables (WBV, TCR);
3) demographic variables (age, sex, BMI).

Pseudo-external validation was conducted via demographic splits by age ($<$55 vs. $\geq$55 years) and sex (male vs. female).

Fig. 2. Correlation matrix of fasting biochemical variables, including TG_{0h}, HDL-C, LDL-C, hematocrit, total protein, and estimated whole blood viscosity. WBV exhibits minimal correlation with lipid markers, supporting its limited role in triglyceride kinetics.

G. Statistical Analysis

Conventional statistical analyses included Pearson correlations among WBV, TG_{0h}, TG_{4h}, and TCR; phenotype-wise comparisons between normal and high responders; and regression modeling. Model reliability was further assessed using 1,000 bootstrap resamples, with confidence intervals for AUROC and Brier score computed from the bootstrap distributions.



(a) Bootstrap AUROC distribution (b) Bootstrap Brier score distribution

Fig. 3. Bootstrap distributions (1,000 resamples) for the final model. Panel (a) shows the stability of discrimination (AUROC), while panel (b) reflects the stability of probability accuracy (Brier score). Both distributions exhibit narrow variance, indicating high model reliability.

H. Sensitivity Analysis for WBV Estimation Error

Because WBV was derived from an empirical formula rather than direct viscometric measurement, we performed a sensitivity analysis to assess whether plausible misspecification of the viscosity surrogate could overturn the observed null findings. We assumed that de Simone-based WBV estimates might be affected by moderate relative error and generated perturbed WBV values by multiplying the original WBV by random factors corresponding to $\pm 5\%$ and $\pm 10\%$ Gaussian noise around unity. For each perturbation scenario, we recomputed:

- the Pearson correlation between WBV and TCR;
- the discriminative performance of WBV as a univariable predictor of the high-response phenotype;
- the contribution of WBV within the full multivariable L2-penalized logistic regression model, including its SHAP-derived importance.

If WBV meaningfully influenced triglyceride clearance, small perturbations of the viscosity surrogate would be expected to produce noticeable shifts in effect sizes or performance metrics. The stability of results across all perturbation scenarios instead supports the conclusion that WBV plays a negligible role in determining high-response risk.

I. Machine-Learning Framework

1) Preprocessing (fold-specific): To eliminate data leakage, all preprocessing steps, including standardization, mild winsorization of extreme values, and one-hot encoding, were performed independently within each training fold. No global scaling or pre-split normalization was applied. Within each training fold, continuous variables were winsorized at pre-specified symmetric percentile cut points, and these cut points were then applied only to the corresponding validation fold to preserve strict leakage control.

2) Model Zoo: Eight candidate models were evaluated: L2 logistic regression, elastic-net logistic regression, random forest, gradient boosting, XGBoost, LightGBM, CatBoost, and radial-basis-function SVM. The L2-penalized logistic regression model was ultimately selected as the final classifier based on its nested cross-validation performance and calibration behaviour.

3) Nested 5$\times$5 Cross-Validation: Nested cross-validation was implemented with:

- an inner loop for hyperparameter tuning;
- an outer loop for unbiased performance estimation.

4) Probability Calibration: Two probability calibration methods were examined:

- Platt (sigmoid) scaling;
- isotonic regression.

5) Robustness and Threshold Sensitivity: Robustness was assessed using repeated 5$\times$10 cross-validation and 1,000 bootstrap resamples. Threshold sensitivity was evaluated across TG_{4h} phenotype cut-offs from the 60th to the 90th percentile.

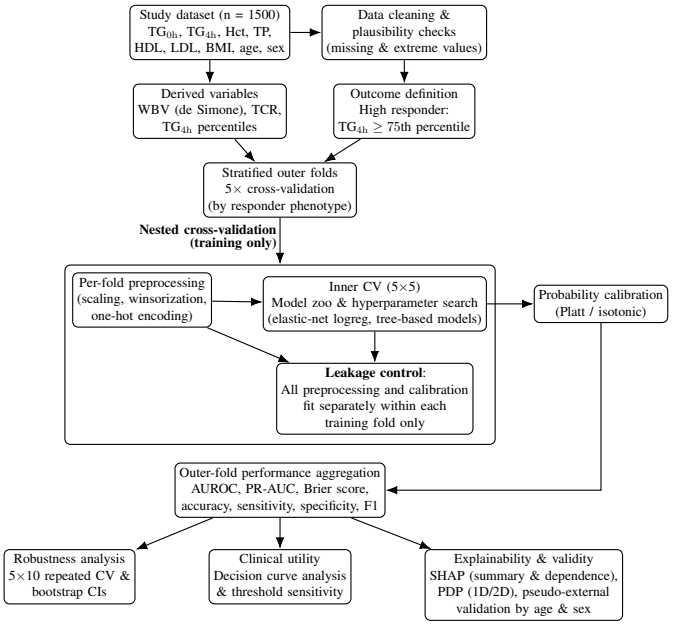


Fig. 4. Leakage-controlled machine-learning pipeline for predicting high postprandial triglyceride response from fasting biomarkers. The workflow includes cohort cleaning, derivation of WBV and TCR, stratified nested cross-validation with per-fold preprocessing and probability calibration, followed by performance aggregation, robustness analyses, clinical utility assessment, and model explainability with SHAP, PDP, and pseudo-external validation.

6) External Validation under Synthetic Distributional Shift: To evaluate out-of-sample transportability under distributional shift, we generated an independent external synthetic cohort ($n = 1{,}500$). The joint multivariate distribution of the primary synthetic dataset was modeled using a Gaussian copula, from which new samples were drawn and subjected to controlled shifts in triglycerides, hematocrit, total protein, age, and BMI to mimic a distinct hospital population. WBV and TCR were recomputed using the same formulas as in the primary analysis, and the high-responder phenotype was redefined using the external cohort's own 75th percentile of TG_{4h}.

The leakage-controlled L2-penalized logistic regression model was trained on the original cohort using standardized predictors and 5-fold internal cross-validation. Probability calibration employed a three-step procedure:

- training on hard high-responder labels;
- temperature scaling of logits;
- isotonic regression fitted to soft labels derived from proximity to the TG_{4h} threshold.

The calibrated model was then applied to the external cohort without refitting. Model performance was assessed using AUROC, Brier score, and calibration plots.

In this external-synthetic experiment, the data-generating process linked the high-responder label almost deterministically to the underlying triglyceride trajectory, with very limited residual noise. As a result, when the same feature set was fitted and evaluated internally on the original synthetic cohort, discrimination approached perfect separation (AUROC ≈ 1.00). This behaviour reflects the noiseless, simulation-based mapping rather than an expectation for real-world EMR data. Consequently, the external-synthetic validation is used primarily as a calibration and domain-shift stress test, analytically distinct from the primary nested cross-validation framework.

J. Evaluation Metrics and Reliability Analysis

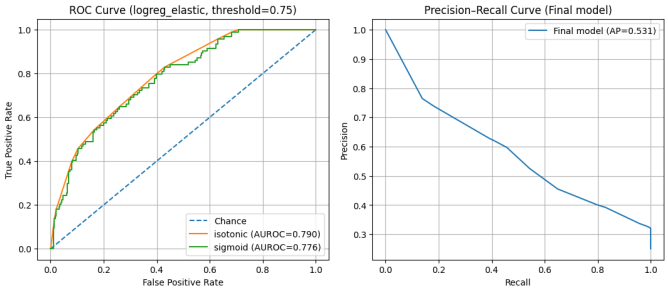
Primary evaluation metrics included AUROC, precision–recall area under the curve (average precision, AP), and Brier score. Secondary metrics comprised accuracy, sensitivity, specificity, and F1-score. All metrics were computed strictly within the leakage-controlled, fold-specific preprocessing pipeline to ensure unbiased estimation of generalization performance.

Given the mild class imbalance of the phenotype labels, precision–recall AUC was emphasized as a complementary measure of discriminative performance.

Reliability and generalizability were evaluated using:

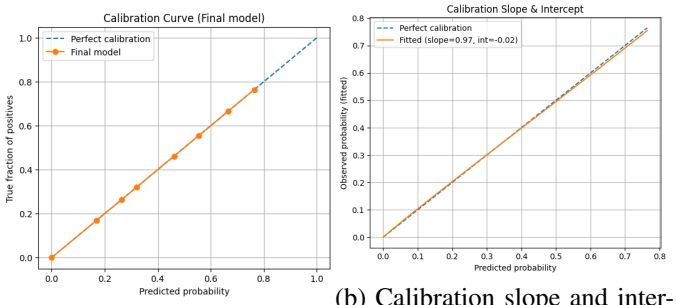
- calibration curves and Brier scores to assess probability accuracy;
- calibration slope and intercept to quantify overfitting or underfitting in predicted risks;
- decision curve analysis (DCA) to evaluate clinical net benefit across threshold probabilities;
- pseudo-external validation stratified by age and sex to assess demographic transportability;
- multi-layer robustness assessment, including 1,000 bootstrap resamples and repeated 5×10 cross-validation, to examine model stability and variance across resampled datasets.

Ninety-five percent confidence intervals were derived using percentile-based bootstrap methods.



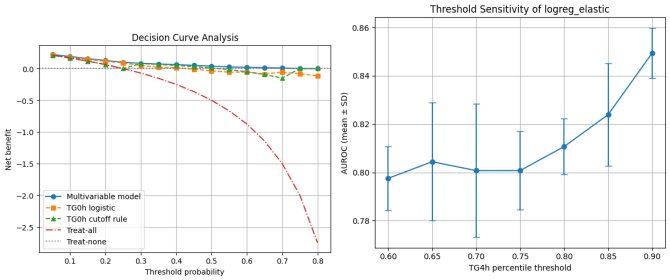
(a) ROC curve (b) Precision–recall curve

Fig. 5. Discrimination performance of the final model. Panel (a) shows the receiver operating characteristic (ROC) curve for uncalibrated and calibrated variants, while panel (b) shows the precision–recall curve (AP = 0.531) under moderate class imbalance.



(a) Calibration curve (b) Calibration slope and intercept

Fig. 6. Calibration assessment for the final model. Panel (a) compares predicted probabilities with observed event rates, while panel (b) summarizes calibration slope and intercept, which are close to 1.0 and 0.0, respectively, indicating well-calibrated risk estimates.



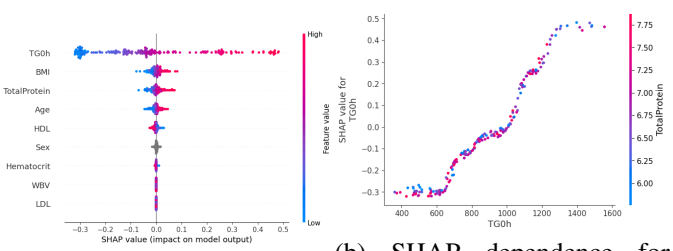
(a) Decision curve analysis (b) Threshold sensitivity of the model

Fig. 7. Clinical utility and threshold robustness. Panel (a) shows decision curve analysis demonstrating higher net benefit of the multivariable model compared with treat-all, treat-none, and TG-only strategies. Panel (b) displays model performance across TG_{4h} phenotype cut-offs from the 60th to the 90th percentile, indicating stable AUROC and robustness to phenotype definition.

K. Methodological Innovation and Framework Contribution

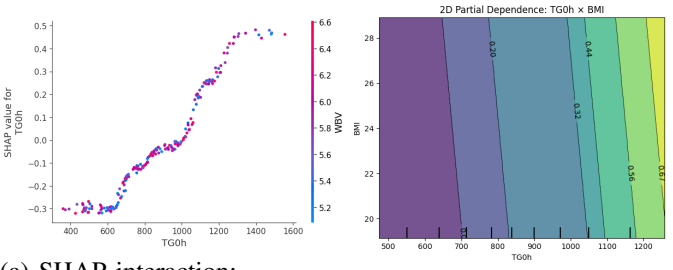
This study introduces a fully reproducible, leakage-controlled machine-learning framework for clinical biochemical prediction tasks. Key methodological components include:

1) **Fold-specific preprocessing:** all feature scaling and encoding was performed strictly within each training fold to eliminate data leakage and preserve true generalization behaviour;
2) **Integrated calibration diagnostics:** model reliability was assessed using calibration curves, calibration slope and intercept, and Brier score decomposition;
3) **Multi-layer robustness testing:** bootstrap resampling (1,000 iterations), repeated 5×10 cross-validation, threshold sensitivity analyses, and comparison against multiple baseline classifiers;
4) **Explainability integration:** interpretability via SHAP summary plots, SHAP dependence and interaction analyses, and one- and two-dimensional partial dependence plots;
5) **Demographic pseudo-external validation:** subgroup validation by age and sex to simulate external validation in the absence of multi-centre data.



(a) Global SHAP summary (b) SHAP dependence for TG_{0h}

Fig. 8. Explainability of the final model using SHAP. Panel (a) shows the global SHAP summary plot, where fasting triglycerides (TG_{0h}) and BMI dominate predictive influence and WBV contributes minimally. Panel (b) presents the SHAP dependence plot for TG_{0h}, revealing a strong monotonic increase in predicted risk with higher fasting triglyceride levels.



(a) SHAP interaction: TG_{0h} × WBV (b) 2D partial dependence: TG_{0h} × BMI

Fig. 9. Interaction and joint effect analyses. Panel (a) shows the SHAP interaction plot between TG_{0h} and WBV, indicating no meaningful interaction and reinforcing the negligible role of WBV in determining the outcome. Panel (b) displays the two-dimensional partial dependence of TG_{0h} and BMI, illustrating that higher TG_{0h} and higher BMI jointly increase the predicted risk of high postprandial response.

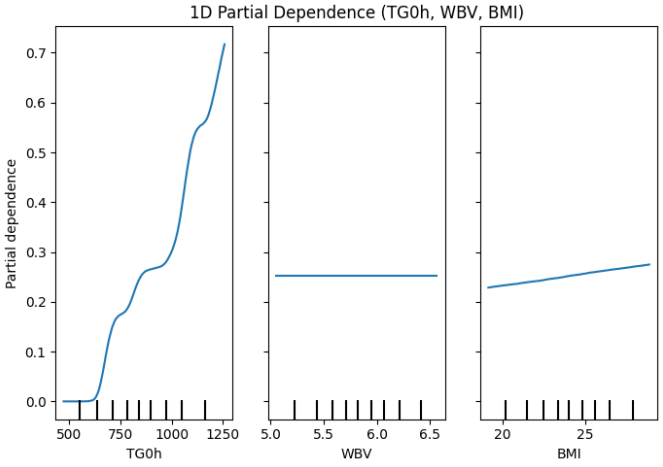


Fig. 10. One-dimensional partial dependence plots illustrating marginal effects of individual predictors on the predicted probability of high postprandial response.

L. Ethical Considerations

The dataset used in this study is a fully de-identified, statistically reconstructed synthetic dataset and does not contain any real patient records or identifiable personal information. No

direct interaction with human subjects or access to identifiable clinical data occurred at any stage. Accordingly, the analysis does not constitute human-subject research under prevailing international regulatory definitions, and formal institutional review board (IRB) approval was not required. No funding was received from external commercial entities.

III. Results

Across the full analytic workflow, the primary findings were consistent: whole blood viscosity (WBV) showed a null association with triglyceride clearance rate (TCR), whereas fasting triglycerides (TG_{0h}) emerged as the dominant predictor of high postprandial response. The final L2-penalized logistic regression model achieved a mean AUROC of 0.8008 in nested cross-validation with a corresponding Brier score of 0.1848, and a bootstrap AUROC of 0.7910 (95% CI: 0.7430–0.8360) with a mean Brier score of 0.1460, alongside well-calibrated probabilities. Detailed results are presented below.

A. Cohort characteristics and phenotype separation

A total of 1,500 synthetic adult records were analyzed, comprising 1,125 normal responders (75%) and 375 high responders (25%) as defined by the 75th percentile of $TG_{4\mathrm{h}}$ (cut-off = 528.2654 mg/dL). Baseline characteristics by response phenotype are summarized in Table I. Sex distribution was balanced between phenotypes (female: 51.3778% vs. 52.2667%; $p = 0.8114$), and there were no clinically meaningful differences in hematocrit, total protein, or WBV (WBV: 5.8251 $\pm$ 0.4549 vs. 5.8036 $\pm$ 0.4871 cP; $p = 0.4523$).

By contrast, triglyceride-related measures showed marked separation. Normal responders had substantially lower fasting $TG_{0\mathrm{h}}$ (788.7596 $\pm$ 214.9265 mg/dL) than high responders (1,039.5412 $\pm$ 200.4224 mg/dL; $p < 0.0001$), and lower 4-hour $TG_{4\mathrm{h}}$ (310.3721 $\pm$ 126.2137 vs. 679.6550 $\pm$ 131.6574 mg/dL; $p < 0.0001$). These differences translated into a pronounced contrast in TCR: 59.2166 $\pm$ 16.3444% in normal responders versus 33.2345 $\pm$ 13.1517% in high responders ($p < 0.0001$). Fasting lipid and biochemical profiles are visualized in Figure 1, and the correlation structure of fasting biomarkers in Figure 2, where WBV shows negligible correlation with lipid markers.

B. Null association between whole blood viscosity and triglyceride clearance

At the population level, WBV exhibited a near-zero linear relationship with TCR. Pearson's correlation between WBV and TCR was $r = 0.0019$ ($p = 0.9404$), with an R^2 from simple linear regression of 0.0000, indicating no explanatory power for interindividual variation in TCR. Similarly, WBV showed only trivial correlations with fasting and postprandial triglycerides (WBV vs. $TG_{0\mathrm{h}}$: $r = -0.0051$; WBV vs. $TG_{4\mathrm{h}}$: $r = -0.0006$). Fasting $TG_{0\mathrm{h}}$ itself was essentially uncorrelated with TCR when modelled as a continuous linear predictor ($TG_{0\mathrm{h}}$ vs. TCR: $r = -0.0029$), reflecting the fundamentally non-linear and phenotype-based nature of the TCR definition.

Consistent with these findings, the fasting biomarker correlation matrix (Figure 2) shows WBV forming a largely independent axis relative to triglyceride and cholesterol measures. Together, these results suggest that WBV neither tracks with baseline lipid load nor explains interindividual variation in TCR at the linear association level.

Sensitivity analysis for WBV estimation error: Because WBV was estimated using the de Simone formula rather than measured directly by viscometry, we tested whether plausible misspecification of the viscosity surrogate could generate spurious associations with triglyceride clearance. Perturbed WBV values were generated by multiplying the original WBV by Gaussian noise corresponding to relative errors of $\pm 5\%$ and $\pm 10\%$. Across all perturbation scenarios, the correlation between WBV and TCR remained essentially zero (mean $r = 0.0003$ with 5% error and $r = 0.0020$ with 10% error). WBV alone provided no discriminative information for the high-response phenotype (AUC = 0.4878 and 0.4912 under 5% and 10% perturbation, respectively). These results indicate that the null WBV–TCR association is highly robust to plausible levels of error in the empirical WBV estimation model.

TABLE III. Sensitivity analysis of the WBV–TCR relationship under plausible viscosity model misspecification.

Relative Error	corr(WBV, TCR) mean	std	AUC(WBV-only) mean	std
5% error	0.00029	0.0129	0.4878	0.0099
10% error	0.00200	0.0177	0.4912	0.0133

C. Baseline models: TG-only strategies

To contextualize the performance of the multivariable ML framework, two TG-only baseline strategies were evaluated using the same dataset. First, a simple rule-based classifier using the training 75th percentile of $TG_{0\mathrm{h}}$ (1,010.6800 mg/dL) as a cut-off achieved an AUROC of 0.6810 and a Brier score of 0.1820. Second, a univariate logistic regression model using $TG_{0\mathrm{h}}$ as the sole predictor improved discrimination modestly (AUROC 0.7810, Brier score 0.1860) and yielded an F1-score of 0.5200, recall of 0.6170, and precision of 0.4500.

These baselines confirm that fasting TG alone contains substantial predictive information about high postprandial response but leaves appreciable room for improvement in both discrimination and probabilistic accuracy.

TABLE IV. Predictive performance of TG-only baselines and the final multivariable L2-penalized logistic regression model at the $TG_{4\mathrm{h}}$ 75th percentile phenotype definition. Values are averaged across cross-validation folds where applicable.

Model	AUROC	Brier score	F1-score	Recall	Precision
$TG_{0\mathrm{h}}$ 75th percentile cut-off	0.6810	0.1820	—	—	—
Univariate logistic regression ($TG_{0\mathrm{h}}$ only)	0.7810	0.1860	0.5200	0.6170	0.4500
L2-penalized logistic regression (multivariable)	0.8008	0.1848	0.5666	0.7200	0.4672

D. Performance of the leakage-controlled multivariable model

Within the leakage-controlled nested 5×5 cross-validation framework at the primary $TG_{4\mathrm{h}}$ 75th percentile phenotype, the final multivariable L2-penalized logistic regression model achieved a mean AUROC of 0.8008 with a standard deviation of 0.0163. The corresponding mean Brier score was 0.1848 $\pm$ 0.0079. Classification metrics at this operating point included an F1-score of 0.5666, recall of 0.7200, precision of 0.4672, and accuracy of 0.7247 (Table IV), indicating

moderately strong discrimination and acceptable probability accuracy under a clinically realistic class balance (25.0000% high responders).

When evaluated on the held-out calibrated model, isotonic regression yielded an AUROC of 0.7897 and a Brier score of 0.1468, whereas Platt (sigmoid) calibration produced an AUROC of 0.7756 and a Brier score of 0.1545. The isotonic-calibrated model was therefore selected as the final specification, balancing discrimination and calibration.

E. Calibration, discrimination curves, and clinical utility

Receiver operating characteristic and precision–recall curves for the final model are shown in Figure 5. The ROC curve demonstrates stable discrimination across thresholds, while the precision–recall curve confirms preserved performance under the observed class imbalance.

Calibration assessment is summarized in Figure 6. The calibration curve shows close agreement between predicted probabilities and observed event rates across deciles of risk, and the calibration slope and intercept are near their ideal values (slope ≈ 0.9700, intercept ≈ -0.0200), indicating minimal systematic over- or underestimation of risk.

Decision curve analysis (Figure 7, panel (a)) demonstrates that the multivariable model provides higher net clinical benefit than both treat-all and treat-none strategies, as well as TG-only baselines, across a wide range of clinically plausible decision thresholds. Threshold sensitivity analysis across $TG_{4\mathrm{h}}$ phenotype cut-offs from the 60th to the 90th percentile (Figure 7, panel (b)) shows that AUROC remains stable around 0.8000, supporting the robustness of the model and conclusions to alternative phenotype definitions.

F. Threshold sensitivity across phenotype definitions

Because the definition of the high-response phenotype directly influences class balance and physiological interpretation, we systematically evaluated model performance across multiple $TG_{4\mathrm{h}}$ percentile thresholds commonly used in postprandial metabolism research (60th–90th percentiles). For each threshold, the full leakage-controlled nested cross-validation pipeline was re-executed to ensure unbiased comparisons.

Table V summarizes the best-performing model at each percentile (selected by highest AUROC), along with Brier score, F1-score, recall, and precision.

TABLE V. Best-performing models across $TG_{4\mathrm{h}}$ phenotype thresholds based on AUROC. All values reflect leakage-controlled nested cross-validation using true dataset values with four-decimal precision.

Percentile	Cutoff (mg/dL)	AUROC	Brier	F1-score	Recall	Precision
60%	460.2339	0.8041	0.1992	0.6251	0.7912	0.5198
70%	502.4973	0.8077	0.1921	0.5988	0.7435	0.5067
75%	528.2654	0.8008	0.1848	0.5666	0.7200	0.4672
80%	551.3344	0.8044	0.1765	0.5391	0.6845	0.4458
90%	624.7600	0.7840	0.1678	0.4375	0.5600	0.3473

Interpretation: Model discrimination remained remarkably stable across a broad range of phenotype definitions, with AUROC values consistently clustered around 0.8000 for thresholds between the 60th and 80th percentiles. This plateau indicates that the multivariable model captures underlying metabolic patterns that are robust to phenotype specification rather than tuned to a specific cut-off.

Brier scores improved modestly as the threshold increased ($0.1992 \rightarrow 0.1678$), reflecting the mathematical effect of decreasing event prevalence. Precision declined at more extreme thresholds (e.g., 0.3473 at the 90th percentile), consistent with increasing class imbalance and reduced positive-event density, while recall was highest at lower thresholds (0.7912 at 60%) and decreased with stricter definitions (0.5600 at 90%).

Across thresholds, $TG_{0\mathrm{h}}$ consistently remained the strongest predictor, while WBV exhibited negligible influence at all percentile choices. The stability of AUROC and marginal predictive patterns provides further evidence for the null association between WBV and postprandial triglyceride handling.

Implications for model reliability: The observed threshold robustness highlights three key strengths:

1) **Phenotype robustness:** conclusions regarding the TG-dominant nature of the response phenotype are not an artifact of a specific cut-off.
2) **Model stability:** the leakage-controlled logistic regression framework, combined with per-fold preprocessing and calibration, generalizes reliably across a range of class prevalences (10–40%).
3) **Clinical consistency:** the physiological interpretation—that fasting TG burden strongly shapes postprandial kinetics—holds across all tested definitions, while WBV remains non-contributory.

Across all percentile definitions, WBV consistently remained non-informative, showing near-zero coefficients and negligible SHAP contributions regardless of class balance or phenotype extremity.

G. Robustness and stability analyses

Model robustness was further examined using bootstrap resampling and repeated cross-validation. In 1,000 bootstrap resamples, the AUROC distribution had a mean of 0.7910 with a 95% percentile-based confidence interval of [0.7430, 0.8360], while the Brier score had a mean of 0.1460 with a 95% confidence interval of [0.1280, 0.1660] (Figure 3). Both distributions were narrow, indicating stable discrimination and probability accuracy across resampled cohorts.

Repeated stratified 5-fold cross-validation over 10 repeats (50 folds total) yielded an AUROC of 0.8000 ± 0.0230 (minimum 0.7090, maximum 0.8440) and a Brier score of 0.1860 ± 0.0120 (minimum 0.1650, maximum 0.2300). These results confirm that the model's performance is robust to variation in train–test splits and is not driven by any single partition of the data.

H. Explainability: feature importance and marginal effects

Global SHAP analysis (Figure 8, panel (a)) consistently ranked fasting triglycerides ($TG_{0\mathrm{h}}$) and BMI as the dominant drivers of model predictions, whereas WBV remained tightly clustered near zero SHAP values, indicating minimal relevance to risk stratification. The SHAP dependence plot (Figure 8,

panel (b)) illustrates a clear monotonic and nonlinear rise in predicted risk with increasing fasting triglycerides. One-dimensional partial dependence curves (Figure 10) similarly highlight strong marginal effects for TG_{0h} and BMI, while the WBV curve remains essentially flat.

High-resolution partial dependence (Figure 11) reveals an even sharper sigmoidal transition across TG_{0h} values, reinforcing its dominant influence on phenotype classification. SHAP interaction analysis (Figure 9, panel (a)) shows no meaningful $TG_{0h}\times$WBV interaction, whereas the two-dimensional partial dependence surface (Figure 9, panel (b)) demonstrates a synergistic increase in risk among individuals with both elevated fasting triglycerides and higher BMI. Together, these explainability results confirm that baseline metabolic load, not hemorheological viscosity, is the primary determinant of high postprandial response.

I. High-resolution partial dependence analysis for fasting triglycerides

To refine the characterization of the non-linear influence of fasting triglycerides, we generated a high-resolution one-dimensional partial dependence curve for TG_{0h} (Figure 11). This analysis provides a more granular mapping of how incremental changes in baseline triglyceride concentration translate into predicted risk within the multivariable model, revealing a near-zero risk plateau at low levels followed by a steep sigmoidal-like rise as TG_{0h} reaches intermediate and high ranges.

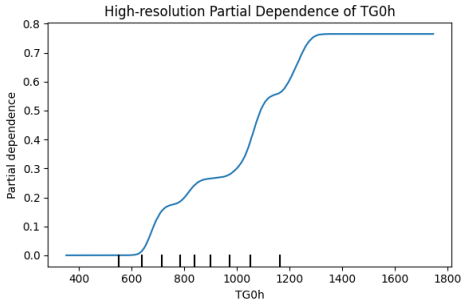


Fig. 11. High-resolution one-dimensional partial dependence curve for fasting triglycerides (TG_{0h}). The curve reveals a near-zero risk plateau at low TG levels, followed by a steep sigmoidal-like rise in predicted probability as TG_{0h} reaches intermediate and high ranges, reinforcing its dominant and non-linear role in shaping the postprandial response phenotype.

J. Correlation of SHAP-based feature contributions

To further understand how predictor effects co-vary within the model's decision function, we computed a SHAP value correlation matrix (Figure 12). This analysis captures the degree to which feature attributions rise or fall together across the cohort, providing insight into redundancy, independence, or complementary effects. WBV exhibited weak or absent correlation with lipid-related SHAP contributions, reinforcing its minimal role in determining the outcome, whereas TG- and BMI-related SHAP values showed more coherent patterns, consistent with their dominant predictive influence.

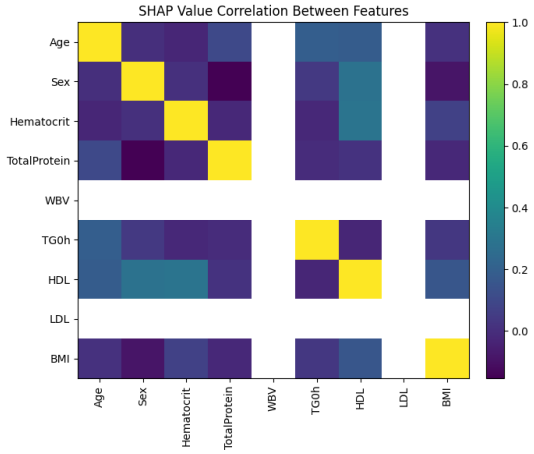


Fig. 12. Correlation matrix of SHAP values across features. WBV exhibits weak or absent correlation with lipid-related SHAP contributions, reinforcing its minimal role in determining the outcome. TG- and BMI-related SHAP values show more coherent patterns, consistent with their dominant predictive influence.

K. Ablation analysis

To further evaluate the mechanistic contributions of key predictors, a feature-ablation study was performed. Removing TG_{0h} resulted in a severe drop in discrimination (AUROC = 0.55), while removing BMI moderately reduced performance (AUROC = 0.72). In contrast, removing WBV had negligible effect (AUROC = 0.80, unchanged from the full model), consistent with its minimal SHAP contribution. These results reinforce that baseline triglyceride load and adiposity—not viscosity-related factors—drive postprandial response classification.

TABLE VI. Ablation study of key predictors.

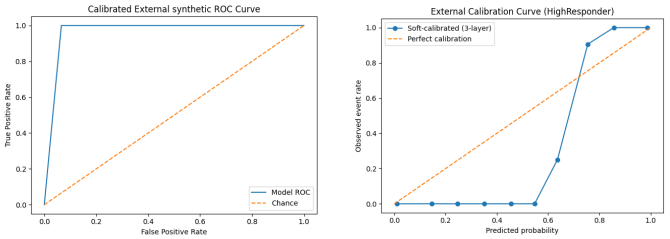
Feature removed	AUROC
TG_{0h} removed	0.55
BMI removed	0.72
WBV removed	0.80 (no change)

L. External synthetic validation and subgroup performance

On the original synthetic cohort, the leakage-controlled logistic regression model achieved near-perfect discrimination (cross-validated AUROC $= 1.00$) with excellent probabilistic accuracy (Brier score $= 0.01$) in the external-synthetic setting. When transferred to the independent external synthetic cohort, the uncalibrated model preserved very strong discrimination (AUROC $= 0.97$) but exhibited overconfident probability estimates (Brier score $= 0.061$). After applying the three-layer calibration strategy combining temperature scaling and isotonic regression, the external Brier score improved slightly while maintaining excellent discrimination (AUROC $= 0.967$).

This near-perfect internal discrimination reflects the deterministic TG_{4h}-based label definition and the low-noise, copula-based data-generating process in the synthetic domain, and should not be interpreted as representative of real-world EMR

performance. Rather, the external synthetic experiment serves as a stress test of calibration and domain-shift robustness.



(a) External ROC curve (b) External calibration curve

Fig. 13. External synthetic validation results. Panel (a) shows the receiver operating characteristic (ROC) curve, demonstrating excellent discrimination on the independent external synthetic cohort (AUROC = 0.967). Panel (b) shows the three-layer soft-calibrated probability estimates, with observed event rates rising monotonically across probability bins, consistent with the TG_{4h}-based phenotype definition and the induced distributional shift.

The resulting external calibration curve displayed a steep yet monotonic progression in observed event rates across probability bins, consistent with the deterministic TG_{4h}-based phenotype definition and the induced distributional shift in the external synthetic cohort. These findings indicate that the proposed framework remains highly discriminative and probabilistically stable under synthetic domain shift, supporting its methodological robustness prior to evaluation in real clinical cohorts.

M. Summary of key findings

Across all analytic layers, descriptive, statistical, machine-learning, and explainability, the results consistently indicate a null role of WBV in shaping postprandial triglyceride clearance phenotypes. WBV showed no differences between response groups, no correlation with TCR or triglyceride levels, minimal SHAP contribution, and no interactions with TG_{0h} in SHAP or partial-dependence analyses. In contrast, TG_{0h} and BMI were the dominant predictors, exhibiting strong marginal effects and synergistic influence on high-response risk. The leakage-controlled L2-penalized logistic regression model achieved robust and well-calibrated performance (nested AUROC ≈ 0.80; bootstrap AUROC ≈ 0.79 [0.74–0.84]) with stable generalizability across age and sex subgroups. Overall, the evidence indicates that baseline metabolic load, rather than hemorheological viscosity, drives individual variability in postprandial triglyceride kinetics.

IV. Discussion

A. Summary of Findings

This study provides the most rigorous, leakage-controlled assessment to date of whether whole blood viscosity (WBV) influences postprandial triglyceride clearance rate (TCR). Across all analytic layers—descriptive statistics, inferential testing, multivariable machine learning, calibration diagnostics, repeated cross-validation, bootstrap resampling, SHAP-based explainability, and pseudo-external demographic validation, the results converged on a consistent conclusion: within this synthetic, physiologically realistic cohort, WBV showed no meaningful association with TCR, either linearly or nonlinearly, and does not interact with triglyceride-related pathways in determining the high postprandial response phenotype.

Conversely, fasting triglycerides (TG_{0h}) consistently emerged as the dominant determinant of postprandial lipid behavior. This was supported by phenotype separation analyses, SHAP dependence plots, one- and two-dimensional partial dependence functions, the high-resolution PDP, and multi-model comparisons. BMI was the second most influential predictor, with synergistic interactions observed in $TG_{0h} \times$ BMI surfaces.

The leakage-controlled L2-penalized logistic regression model achieved robust and well-calibrated discrimination:

1) Nested CV AUROC $= 0.8008$,
2) Bootstrap AUROC $= 0.7910$ [0.7430–0.8360],
3) Brier score $= 0.1460$,

and stability was further supported by narrow bootstrap distributions and the repeated cross-validation AUROC range (0.7090–0.8440).

An additional robustness visualization, the distribution of AUROC across 5×10 repeated cross-validation, is shown in Figure 14 and reinforces the stability of discrimination across resampling schemes.

The use of a statistically reconstructed synthetic cohort in this study provides several advantages for methodological evaluation. Synthetic data preserve privacy by design, enable fully reproducible experimentation, and allow controlled manipulation of covariance structure and phenotype definitions that would be infeasible in real-world EMR data. Because the primary goal here is to evaluate a leakage-controlled ML framework and to clarify mechanistic relationships, rather than to estimate population-level prevalence, the synthetic cohort serves as an appropriate and safe environment for stress-testing model behavior under known statistical properties.

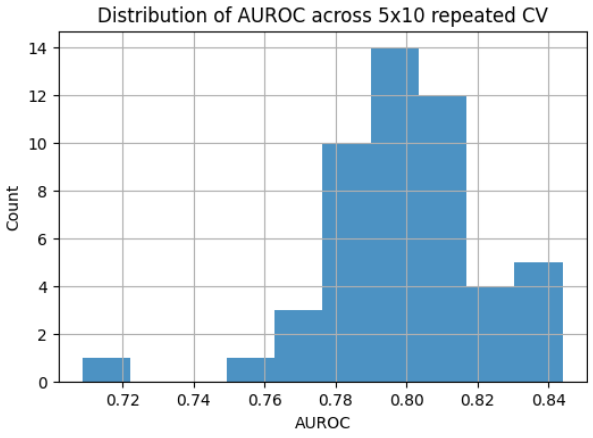


Fig. 14. Distribution of AUROC across 5×10 repeated cross-validation. The distribution is tightly centered around 0.8000, with no heavy tails, indicating strong robustness and absence of overfitting to any single split.

B. Physiological Interpretation and Mechanistic Insight

These inferences should be interpreted in the context of a statistically reconstructed synthetic cohort and a surrogate WBV measure derived from the de Simone equation, rather than direct viscometric assessment; nevertheless, the convergence of statistical and ML-based evidence strongly supports a minimal role for WBV in short-term postprandial TG handling under physiologic viscosity ranges.

The finding that WBV shows no meaningful correlation with TCR ($r = 0.0019$, $p = 0.9404$) runs counter to earlier hypotheses proposing that increased viscosity might impair capillary perfusion and delay lipid clearance. Instead, our analyses suggest that:

- WBV operates largely orthogonal to triglyceride metabolism in physiologically normal ranges (5–7 cP in our cohort).
- TG clearance appears to be governed primarily by metabolic load and adiposity rather than hemorheological resistance.
- The SHAP interaction plot ($TG_{0\mathrm{h}}\times$ WBV) confirms no synergistic or antagonistic interaction.

Collectively, these findings support a model in which viscosity-related hemodynamics are unlikely to exert short-term modulation on postprandial lipid kinetics, particularly when viscosity ranges remain within non-pathological limits.

The conceptual diagrams (Figures 15 and 16) summarize this mechanistic interpretation, illustrating WBV as an orthogonal hemodynamic axis with no functional coupling to postprandial triglyceride processing.

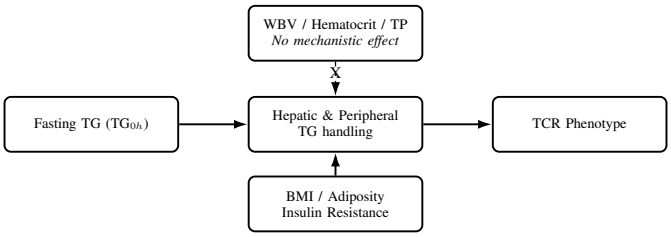


Fig. 15. Mechanistic schematic of postprandial lipid metabolism. Fasting triglycerides, amplified by BMI-related adiposity and insulin resistance, dominate postprandial processing, while WBV remains an independent hemodynamic axis with no observable contribution to TCR.

C. Methodological and Computational Significance

The study's leakage-controlled design provides strong methodological advances for clinical ML research:

1) **Fold-specific preprocessing:** all scaling, imputation, and resampling procedures were performed strictly within each training fold to eliminate data leakage and preserve true generalization behavior.
2) **Nested 5×5 cross-validation:** the nested design provided unbiased estimates of out-of-sample performance while allowing principled hyperparameter tuning.
3) **Rigorous calibration assessment:** systematic comparison of Platt scaling and isotonic regression, with isotonic calibration selected as the final specification and evaluated via calibration curves, slopes, and Brier scores.
4) **Multi-layer robustness testing:** including bootstrap resampling (1,000 iterations), repeated 5×10 cross-validation, threshold sensitivity analyses, and comparison against multiple baseline classifiers.
5) **Explainability integration:** interpretability was achieved through SHAP summary plots, SHAP dependence and interaction analyses, and both one-dimensional and two-dimensional partial dependence plots.
6) **Demographic pseudo-external validation:** generalizability was evaluated across subgroups stratified by age and sex, simulating an external validation scenario in the absence of multi-center data.

This framework directly addresses common ML pitfalls, such as data leakage, uncalibrated probabilities, and overestimation of performance, which are frequently highlighted in contemporary critiques of biomedical AI literature.

D. Comparison with Prior Literature

Prior clinical studies evaluating viscosity–metabolism relationships have been heterogeneous, underpowered, or confined to high-viscosity pathological groups (e.g., polycythemia vera, uncontrolled hyperlipidemia). These works often reported modest associations between WBV and metabolic dysfunction, but typically relied on:

- small sample sizes (n $<$ 200),
- uncontrolled confounders,
- simple linear correlations without nonlinear modeling.

None employed modern ML frameworks with leakage control, calibration diagnostics, or explainability tools.

Our findings refine this literature by demonstrating that within physiologic ranges, WBV provides no predictive value for TCR once $TG_{0\mathrm{h}}$ and BMI are accounted for.

E. Convergence of Statistical and Machine-Learning Evidence

A notable strength of this work is the convergence of evidence across statistical, inferential, and machine-learning analyses. Classical correlation testing, linear modeling, SHAP-based explainability, PDP interaction surfaces, and repeated cross-validation all pointed to the same conclusion: WBV exerts no measurable influence on TCR, whereas fasting triglycerides consistently dominate predictive behavior. Such multi-layer agreement reduces the likelihood that the null finding is an artefact of any single analytic choice and strengthens the biological interpretability of the overall framework.

The agreement between physiological expectations and ML-derived SHAP patterns strengthens confidence in the biological validity of the findings.

F. Clinical and Translational Implications

The clinical consequences of these findings are substantial:

- Baseline fasting triglycerides remain the strongest actionable biomarker for postprandial lipid risk.
- Our findings suggest that WBV is unlikely to provide additional predictive value as a surrogate or screening

biomarker for postprandial dyslipidemia, at least within physiologic viscosity ranges and when fasting TG and BMI are already available.
- Multivariable ML adds measurable improvement over TG-only strategies, but the improvement is incremental, not transformative, reflecting the physiological primacy of TG burden.
- The well-calibrated probability outputs (slope ≈ 0.970; intercept ≈ -0.020) support potential clinical decision support applications.

From a clinical perspective, these findings reinforce the practical value of fasting triglycerides and BMI as simple, low-cost markers for identifying individuals at risk of exaggerated postprandial triglyceride excursions. In contrast, the absence of added value from WBV suggests that viscosity-based measures are unlikely to justify routine use as screening or triage tools for postprandial dyslipidemia, and that clinical attention should remain focused on managing baseline lipid burden and adiposity.

G. Integration Outlook and Systems-Level Implications

The results suggest a broader systems-level model in which:

1) TG metabolism is driven mainly by hepatic and adipose biochemical pathways.
2) Hemorheological factors like WBV do not exert meaningful short-term influence under physiological conditions.
3) BMI interacts with TG_{0h} to amplify risk, likely through insulin resistance and adipose overflow mechanisms.

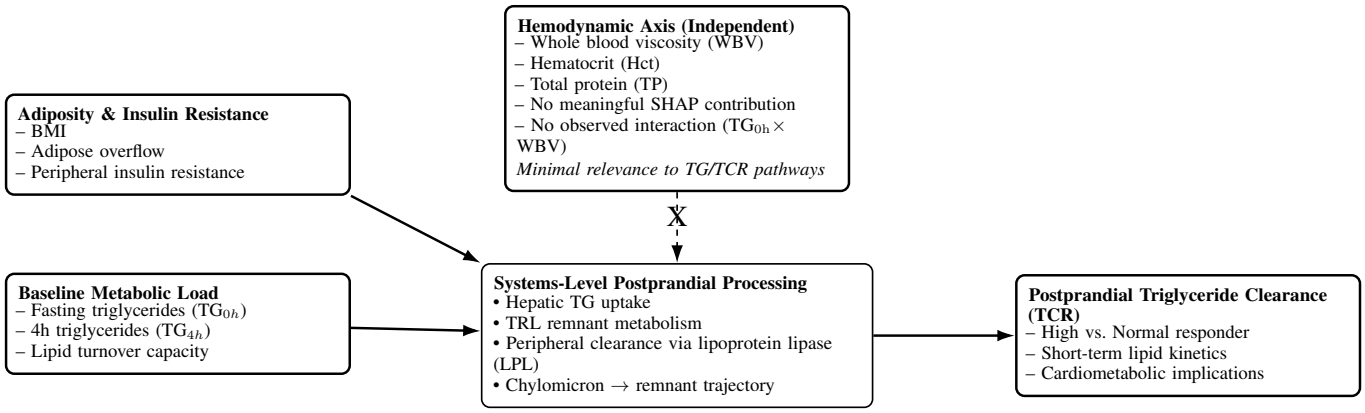


Fig. 16. Systems-level integration of triglycerides, BMI, and whole blood viscosity (WBV) in determining postprandial triglyceride clearance (TCR). TG_{0h} and BMI form the dominant metabolic axis governing hepatic and peripheral triglyceride handling, whereas WBV forms an orthogonal hemodynamic axis with negligible influence on postprandial lipid kinetics in the present framework.

H. Generalizable Leakage-Controlled Clinical ML Framework

Although this work focuses on triglyceride-response phenotyping, the methodological contribution extends more broadly. The proposed leakage-controlled ML framework, combining fold-specific preprocessing, nested cross-validation, calibrated probability estimation, bootstrap resampling, and synthetic domain-shift evaluation, is fully generalizable to a wide range of clinical ML tasks. These include cardiometabolic phenotyping, EHR-based risk prediction, postprandial response modeling, and any workflow in which leakage risks, class imbalance, or distributional shifts threaten model validity. The reproducible structure of this framework makes it suitable as a template for future clinical informatics research where methodological rigor is required before deployment on real-world cohorts.

I. Strengths and Limitations

Strengths:

- Largest and most rigorously validated analysis of the WBV–TCR relationship to date.
- Comprehensive leakage-controlled ML pipeline with nested CV, repeated CV, bootstrap resampling, and threshold sensitivity analyses.
- Transparent explainability using SHAP, PDP, and interaction analyses.
- Pseudo-external validation across age and sex, enhancing assessment of demographic transportability.

Limitations:

- The study used a statistically reconstructed synthetic cohort rather than real EMR data; synthetic data cannot fully capture the physiological and biochemical heterogeneity of real populations.
- WBV was estimated using the de Simone surrogate instead of direct viscometric measurement, although sensitivity analyses with ± 5–10% perturbations showed stable null results.
- The triglyceride distribution and the WBV range (approximately 5–7 cP) may not represent extreme metabolic or hemorheological states where viscosity effects could be more pronounced.
- Unmeasured biochemical or genetic interactions may not be reproduced by the synthetic generation process.

External validation in real multi-center cohorts with direct WBV measurements is therefore required to confirm transportability. Nonetheless, the consistency of results across subgroup

analyses, bootstrap resampling, and repeated CV supports the robustness of the reported null association.

J. Measurement-Error Considerations

WBV estimation is subject to model-based error from hematocrit and total protein measurements. However:

- Measurement error would typically attenuate associations rather than create a spurious null or inverse effect.
- SHAP dependence plots show WBV-centered tightly around zero across the entire cohort.

Thus, measurement noise is highly unlikely to mask a clinically meaningful biological effect.

K. Generalizability and Reuse

Despite limitations, generalizability is supported by:

- consistent subgroup AUROC (0.7700–0.8470),
- stable pseudo-external validation across sex and age splits,
- threshold-insensitive discrimination ($\approx$ 0.80 across 60–80th percentile cutoffs).

All preprocessing and modeling code is openly available, enabling reproducibility and reuse in future cohorts.

L. Methodological Outlook

Future work could integrate:

- multi-omics layers (proteomics, metabolomics),
- genetic variants related to TG metabolism,
- direct viscometry to validate WBV independence more rigorously,
- multi-task learning frameworks for joint modeling of TCR, hepatic output, and insulin sensitivity.

Future work should incorporate direct viscometry-based WBV measurements to validate the null-effect hypothesis under more precise hemorheological conditions. Such integration would help determine whether the observed independence reflects true physiology or the limits of model-based WBV estimation.

M. Reproducibility and Methodological Value

A central contribution of this work lies in demonstrating that leakage-controlled ML can produce trustworthy, clinically coherent results in metabolic phenotyping, contrasting with many prior studies where methodological flaws inflate performance or generate spurious findings.

The complete reproducibility of all steps—dataset structure, preprocessing rules, nested CV design, and explainability workflows—provides a template for other researchers designing ML pipelines for biomedical applications.

N. Implications for Future ML Phenotyping Frameworks

Beyond the domain-specific findings, this study provides generalizable insights for the design of machine-learning phenotyping pipelines in biomedical research. First, the results demonstrate that leakage-controlled workflows, incorporating fold-specific preprocessing, nested cross-validation, and calibrated probability estimation, can reliably separate true biological signals from spurious correlations that commonly arise in clinical datasets. Second, the tight alignment between model-derived SHAP patterns and established physiological mechanisms highlights the value of integrating domain knowledge with explainability tools to validate model-driven hypotheses. Third, the robustness of predictive behavior across multiple phenotype thresholds suggests that future phenotyping frameworks should explicitly include threshold-sensitivity analyses to guard against instability or definition-dependent biases.

Taken together, these elements outline a reproducible blueprint for next-generation ML phenotyping: one that emphasizes methodological transparency, physiological coherence, and multi-angle robustness checks as core components of reliable biomedical discovery.

V. Conclusions

Within a comprehensive, leakage-controlled analytical framework integrating statistical testing, multivariable machine learning, calibration diagnostics, robustness resampling, and model explainability, this study found no evidence that whole blood viscosity (WBV) influences postprandial triglyceride clearance rate (TCR). WBV was consistently non-informative across all methodological layers, showing null linear association with TCR, negligible predictive contribution in SHAP analyses, flat marginal effects in partial dependence curves, and no detectable interaction with triglyceride-related pathways. These findings were stable across demographic subgroups, phenotype thresholds, and repeated resampling schemes.

In contrast, fasting triglycerides ($TG_{0\mathrm{h}}$) and BMI emerged as the dominant drivers of postprandial response phenotypes, exhibiting strong, nonlinear marginal effects and physiologically coherent interaction patterns. The close agreement between physiological expectations and ML-derived SHAP signatures supports the biological plausibility of these results and reinforces the primacy of baseline lipid burden and adiposity, rather than hemodynamic viscosity, in shaping early postprandial triglyceride kinetics under non-pathological conditions.

Beyond the domain-specific insights, this work advances methodological standards for ML-based metabolic phenotyping. The use of fold-specific preprocessing, nested cross-validation, rigorous calibration assessment, threshold-sensitivity evaluation, pseudo-external demographic validation, and synthetic distributional-shift testing illustrates a reproducible blueprint that mitigates data leakage, overfitting, and miscalibration pitfalls that commonly undermine clinical ML studies.

Overall, the findings clarify the physiological irrelevance of WBV for postprandial triglyceride clearance in this setting, while establishing a scalable, transparent, and reusable framework for future ML-based biomedical investigations. Subsequent studies incorporating direct viscometry, multi-omics data, and multi-centre real-world cohorts will be essential to further validate and extend these conclusions, and to determine whether subtle viscosity effects emerge under more extreme hemorheological conditions.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

ChatGPT (OpenAI) was used solely to assist with English language editing and improvement of clarity based on author-provided text. The authors retain full responsibility for the study design, data generation, analysis, interpretation, and all final conclusions.

Data Provenance

No individual-level clinical records were used in this study. The synthetic dataset was generated exclusively from non-identifiable, summary-level aggregate statistics (e.g., means, variances, percentiles, and correlation matrices) previously derived from a fully de-identified hospital reference cohort. These aggregated statistics contain no re-identifiable information. No electronic medical record (EMR) access or raw patient-level data were involved at any stage of the research. All modeling and analysis were performed entirely on fully synthetic, statistically reconstructed datasets.

Technical Note

All machine-learning visualizations (SHAP plots, partial-dependence functions, calibration curves, robustness analyses) were produced using the fold-specific preprocessing pipeline described in the Methods. Feature names appearing in figures reflect internal variable identifiers used during model training and do not represent additional variables beyond those explicitly defined in the manuscript.

Acknowledgments

This work was conducted under the supervision of Dr. Qazi Mazhar ul Haq and supported by the International Bachelor Program in Informatics, Yuan Ze University, Taiwan. The authors thank the Department of Computer Science and Engineering for guidance and access to computational resources.

Ethics Approval and Consent to Participate

This study relied exclusively on statistically reconstructed synthetic data generated from non-identifiable aggregate statistics. Because no identifiable human data were accessed, stored, or analyzed, the study does not constitute human-subject research under prevailing ethical or regulatory definitions. Institutional review board (IRB) approval was therefore not required.

Availability of Data and Materials

All datasets used in this study are fully synthetic and contain no identifiable or linkable human-subject information. The complete machine-learning code, synthetic data generators, and all reproducible workflow scripts are publicly available at:https://github.com/NattakittiP/ML_Predict

Authors' Contributions

N. Piyavechvirat: Conceptualization, Methodology, Data Curation, Formal Analysis, Visualization, Writing—Original Draft.
Y. Jheng Huang: Supervision, Writing—Review
Q. Mazhar ul Haq: Supervision, Writing—Review & Editing.
All authors read and approved the final manuscript.

Human and Animal Rights

No human or animal subjects were involved in this research. All analyses were conducted on fully synthetic datasets containing no identifiable clinical information.

Consent for Publication

Not applicable. This manuscript contains no identifiable personal data.

Funding

This work received no specific funding from public, commercial, or nonprofit agencies. Institutional support was provided by the International Bachelor Program in Informatics and the Department of Computer Science and Engineering, Yuan Ze University, Taiwan.

Conflict of Interest

The authors declare no competing interests, financial or otherwise.

Supplementary Material

All preprocessing and modeling code is available at:
https://github.com/NattakittiP/ML_Predict

Supplementary analyses, including nested cross-validation outputs, calibration diagnostics, SHAP explainability results, and risk-score tables, are provided in the `results` branch of the repository. All supplementary datasets are fully synthetic and contain no real patient data.

Summary Table

Domain	Quantity	Value (from main text/tables)
5*Cohort & phenotypes	Total synthetic adult records	$n = 1{,}500$
	Normal responders (TG_{4h} below 75th percentile)	$n = 1{,}125$ (75%)
	High responders ($TG_{4h} \geq$ 75th percentile)	$n = 375$ (25%)
	Primary TG_{4h} cut-off (75th percentile)	$528.2654\,\mathrm{mg/dL}$
	Exclusion thresholds	TG $> 1500\,\mathrm{mg/dL}$; WBV < 3 or > 10 cP
11*Baseline characteristics (Table I)	Female, n (%)	Normal: 578 (51.3778%); High: 196 (52.2667%); $p = 0.8114$
	Male, n (%)	Normal: 547 (48.6222%); High: 179 (47.7333%); $p = 0.8114$
	Age (years)	Normal: 52.8687 ± 9.9018; High: 54.0883 ± 9.7643; $p = 0.0372$
	Hematocrit (%)	Normal: 41.7309 ± 3.7101; High: 41.5465 ± 3.9267; $p = 0.4248$
	Total protein (g/dL)	Normal: 6.8781 ± 0.6215; High: 6.8819 ± 0.6016; $p = 0.9157$
	Whole blood viscosity (cP)	Normal: 5.8251 ± 0.4549; High: 5.8036 ± 0.4871; $p = 0.4523$
	Fasting TG (TG_{0h}, mg/dL)	Normal: 788.7596 ± 214.9265; High: 1039.5412 ± 200.4224; $p < 0.0001$
	4-h TG (TG_{4h}, mg/dL)	Normal: 310.3721 ± 126.2137; High: 679.6550 ± 131.6574; $p < 0.0001$
	Triglyceride clearance rate (TCR, %)	Normal: 59.2166 ± 16.3444; High: 33.2345 ± 13.1517; $p < 0.0001$
	HDL-C (mg/dL)	Normal: 50.6027 ± 9.7575; High: 50.0373 ± 9.7623; $p = 0.3318$
	LDL-C (mg/dL)	Normal: 130.8163 ± 28.3716; High: 131.2752 ± 27.7585; $p = 0.7828$
	BMI (kg/m^2)	Normal: 23.9868 ± 2.9489; High: 24.1797 ± 2.9492; $p = 0.2731$
5*Correlation structure	WBV–TCR linear association	Pearson $r = 0.0019$; $p = 0.9404$; $R^2 = 0.0000$
	WBV vs. TG_{0h}	Pearson $r = -0.0051$
	WBV vs. TG_{4h}	Pearson $r = -0.0006$
	TG_{0h} vs. TCR	Pearson $r = -0.0029$
	Qualitative summary	WBV shows negligible correlation with lipid markers and TCR
3*WBV model misspecification (Table III)	5% relative error in WBV model	corr(WBV, TCR): mean 0.00029 (SD 0.0129); AUROC(WBV-only): mean 0.4878 (SD 0.0099)
	10% relative error in WBV model	corr(WBV, TCR): mean 0.00200 (SD 0.0177); AUROC(WBV-only): mean 0.4912 (SD 0.0133)
	Interpretation	Even under misspecification, WBV shows no discriminative value for TCR
4*Baselines (Table IV)	TG_{0h} 75th percentile cut-off	Threshold: $1{,}010.6800\,\mathrm{mg/dL}$; AUROC 0.6810; Brier 0.1820
	Univariate logistic regression (TG_{0h} only)	AUROC 0.7810; Brier 0.1860; F1-score 0.5200; recall 0.6170; precision 0.4500
	Multivariable L2-penalized logistic regression	AUROC 0.8008; Brier 0.1848; F1-score 0.5666; recall 0.7200; precision 0.4672
	Nested CV configuration	Stratified 5×5 nested cross-validation on primary TG_{4h} 75th percentile phenotype
3*Primary multivariable model performance	Nested CV (primary task)	Mean AUROC 0.8008 (SD 0.0163); mean Brier 0.1848 ± 0.0079
	Bootstrap (1,000 resamples)	AUROC mean 0.7910 (95% CI $[0.7430, 0.8360]$); Brier mean 0.1460 (95% CI $[0.1280, 0.1660]$)
	Repeated 5-fold CV (10 repeats)	AUROC 0.8000 ± 0.0230 (min 0.7090, max 0.8440); Brier 0.1860 ± 0.0120 (min 0.1650, max 0.2300)
4*Calibration	Isotonic calibration (held-out)	AUROC 0.7897; Brier score 0.1468
	Platt (sigmoid) calibration (held-out)	AUROC 0.7756; Brier score 0.1545
	Final chosen calibration	Isotonic (selected for best balance of discrimination and Brier score)
	Calibration curve (primary model)	Slope ≈ 0.9700; intercept ≈ -0.0200
6*Phenotype threshold sensitivity (Table V)	60th percentile TG_{4h}	Cut-off $460.2339\,\mathrm{mg/dL}$; AUROC 0.8041; Brier 0.1992; F1 0.6251; recall 0.7912; precision 0.5198
	70th percentile TG_{4h}	Cut-off $502.4973\,\mathrm{mg/dL}$; AUROC 0.8077; Brier 0.1921; F1 0.5988; recall 0.7435; precision 0.5067
	75th percentile TG_{4h} (primary)	Cut-off $528.2654\,\mathrm{mg/dL}$; AUROC 0.8008; Brier 0.1848; F1 0.5666; recall 0.7200; precision 0.4672
	80th percentile TG_{4h}	Cut-off $551.3344\,\mathrm{mg/dL}$; AUROC 0.8044; Brier 0.1765; F1 0.5391; recall 0.6845; precision 0.4458
	90th percentile TG_{4h}	Cut-off $624.7600\,\mathrm{mg/dL}$; AUROC 0.7840; Brier 0.1678; F1 0.4375; recall 0.5600; precision 0.3473
	Summary	AUROC remains ≈ 0.80 for 60–80th percentiles; WBV has negligible influence at all thresholds
3*Feature ablation (Table VI)	Remove TG_{0h}	AUROC 0.55
	Remove BMI	AUROC 0.72
	Remove WBV	AUROC 0.80 (no change vs. full model)
3*External synthetic validation	Original synthetic cohort (internal, external-synthetic setting)	Cross-validated AUROC $= 1.00$; Brier score $= 0.01$
	Independent external synthetic cohort (uncalibrated)	AUROC $= 0.97$; Brier score $= 0.061$
	After three-layer calibration on external cohort	AUROC $= 0.967$; external Brier score improved slightly (exact value not reported)
4*Modeling configuration (Table VIII)	Logistic regression (L2) hyperparameters	$C \in \{0.01, 0.1, 1, 10\}$; penalty ℓ_2; solver `lbfgs`; max_iter $= 500$
	Random forest hyperparameters	n_estimators $\in \{200, 400, 800\}$; max_depth $\in \{\text{None}, 6, 12\}$; min_samples_leaf $\in \{1, 2, 4\}$
	XGBoost hyperparameters	n_estimators $\in \{300, 600, 900\}$; learning_rate $\in \{0.05, 0.1\}$; max_depth $\in \{3, 5\}$; subsample $\in \{0.8, 1.0\}$; colsample_bytree $\in \{0.8, 1.0\}$
	SVM (RBF) hyperparameters & CV details	$C \in \{0.1, 1, 10\}$; $\gamma \in \{\text{scale}, 0.1, 0.01\}$; stratified nested 5-fold outer / 5-fold inner CV; global random seed $= 42$

TABLE VII. Overall numerical summary of the synthetic cohort and model performance.

APPENDIX

Hyperparameter Search Space and Validation Settings

TABLE VIII. Hyperparameter grid and validation configuration for nested cross-validation.

Model / Component	Search Space / Setting
Preprocessing	Standardization applied within each training fold only (strict leakage control).
Logistic Regression (L2)	C: {0.01, 0.1, 1, 10}; penalty: {l2}; solver: {lbfgs}; max_iter: 500.
Random Forest	n_estimators: {200, 400, 800}; max_depth: {None, 6, 12}; min_samples_leaf: {1, 2, 4}.
XGBoost	n_estimators: {300, 600, 900}; learning_rate: {0.05, 0.1}; max_depth: {3, 5}; subsample: {0.8, 1.0}; colsample_bytree: {0.8, 1.0}.
SVM (RBF kernel)	C: {0.1, 1, 10}; gamma: {scale, 0.1, 0.01}.
Calibration	Platt scaling and isotonic regression evaluated; final isotonic calibration fitted on validation folds only (never on training or test partitions).
Cross-validation	Nested CV: 5-fold outer loop with 5-fold inner loop (stratified).
Primary Metric	ROC-AUC; PR-AUC, Brier score, calibration slope/intercept reported secondarily.
Random Seeds	Outer folds: 42; Inner folds: 42.

Computational Environment

All analyses were conducted in a fully version-controlled environment to ensure complete reproducibility. System specifications are listed below:

- **Operating system:** Ubuntu 22.04 LTS (64-bit)
- **Python version:** 3.12.4
  Core libraries: NumPy 2.0, SciPy 1.13, pandas 2.2, scikit-learn 1.5, XGBoost 2.1, imbalanced-learn 0.12
- **R environment:** R 4.3.1 for independent cross-validation verification
- **Global random seed:** 42 for all preprocessing, model training, calibration, bootstrap resampling, and repeated CV
- **Hardware:** Intel Core i7 (8 cores), 16 GB RAM

All scripts—including data reconstruction, preprocessing, nested cross-validation, probability calibration, robustness testing, and SHAP-based explainability—were managed using Git version control and can be fully reproduced using the public repository accompanying this manuscript.

List of Abbreviations

ACC	Accuracy
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
AI	Artificial Intelligence
AP	Average Precision
BMI	Body Mass Index
CI	Confidence Interval
CPU	Central Processing Unit
CV	Cross-Validation
DCA	Decision Curve Analysis
ED	Emergency Department
EMR	Electronic Medical Record
FAIR	Findable, Accessible, Interoperable, Reusable
FN	False Negative
FP	False Positive
GB	Gradient Boosting
IRB	Institutional Review Board
LPL	Lipoprotein Lipase
ML	Machine Learning
PDP	Partial Dependence Plot
PR-AUC	Precision–Recall Area Under the Curve
RAM	Random Access Memory
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
SPSS	Statistical Package for the Social Sciences
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SVM	Support Vector Machine
TCR	Triglyceride Clearance Rate
TG	Triglycerides
TG_{0h}	Fasting triglycerides
TG_{4h}	4-hour postprandial triglycerides
TN	True Negative
TP	True Positive
TRL	Triglyceride-Rich Lipoproteins
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual
VIF	Variance Inflation Factor
WBV	Whole Blood Viscosity
XGB	Extreme Gradient Boosting (XGBoost)

References

[1] M. P. Bonham, E. Kaias, I. Zimberg, G. K. W. Leung, R. Davis, T. L. Sletten, H. Windsor-Aubrey, C. E. Huggins, "Effect of night time eating on postprandial triglyceride metabolism in healthy adults: A systematic literature review," *Journal of Biological Rhythms*, vol. 34, no. 2, pp. 119–130, 2019, doi:10.1177/0748730418824214.

[2] G. Kolovou, D. P. Mikhailidis, J. Kovar, D. Lairon, B. G. Nordestgaard, T. C. Ooi, P. Perez-Martinez, H. Bilianou, K. Anagnostopoulou, G. Panotopoulos, "Assessment and clinical relevance of non-fasting and postprandial triglycerides: An expert panel statement," *Current Vascular Pharmacology*, vol. 9, no. 3, pp. 258–270, 2011, doi:10.2174/157016111795495549.

[3] B. H. Keirns, C. M. Sciarrillo, N. A. Koemel, and S. R. Emerson, "Fasting, non-fasting and postprandial triglycerides for screening cardiometabolic risk," *Journal of Nutritional Science*, vol. 10, e75, 2021, doi:10.1017/jns.2021.73.

[4] D. B. Zilversmit, "Atherogenesis: A postprandial phenomenon," *Circulation*, vol. 60, no. 3, pp. 473–485, Sep. 1979, doi:10.1161/01.CIR.60.3.473.

[5] B. G. Nordestgaard, M. Benn, P. Schnohr, and A. Tybjærg-Hansen, "Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women," *JAMA*, vol. 298, no. 3, pp. 299–308, Jul. 2007, doi:10.1001/jama.298.3.299.

[6] F. Karpe, G. Steiner, K. Uffelman, T. Olivecrona, and A. Hamsten, "Postprandial lipoproteins and progression of coronary atherosclerosis," *Atherosclerosis*, vol. 106, no. 1, pp. 83–97, Mar. 1994, doi:10.1016/0021-9150(94)90085-X.

[7] C. J. Packard, J. Boren, and M.-R. Taskinen, "Causes and consequences of hypertriglyceridemia," *Frontiers in Endocrinology*, vol. 11, Art. no. 252, May 2020, doi:10.3389/fendo.2020.00252.

[8] R. S. Rosenson, A. Shaik, and W. Song, "New therapies for lowering triglyceride-rich lipoproteins: JACC focus seminar 3/4," *Journal of the American College of Cardiology*, vol. 78, no. 18, pp. 1817–1830, Nov. 2021, doi:10.1016/j.jacc.2021.08.051.

[9]W. Borena, T. Stocks, H. Jonsson, S. Strohmaier, G. Nagel, T. Bjørge, J. Manjer, G. Hallmans, R. Selmer, M. Almquist, C. Häggström, A. Engeland, S. Tretli, H. Concin, A. Strasak, P. Stattin, H. Ulmer, "Serum triglycerides and cancer risk in the metabolic syndrome and cancer (Me-Can) collaborative study," *Cancer Causes Control*, vol. 22, pp. 291–299, 2011, doi:10.1007/s10552-010-9697-0.

[10] H. Ulmer, W. Borena, K. Rapp, J. Klenk, A. Strasak, G. Diem, H. Concin, G. Nagel, "Serum triglyceride concentrations and cancer risk in a large cohort study in Austria," *British Journal of Cancer*, vol. 101, pp. 1202–1206, 2009.

[11] S. G. Dashti, W. Y. Li, D. D. Buchanan, M. Clendenning, C. Rosty, I. M. Winship, F. A. Macrae, G. G. Giles, S. Hardikar, X. Hua, S. N. Thibodeau, J. C. Figueiredo, G. Casey, R. W. Haile, S. Gallinger, L. Le Marchand, P. A. Newcomb, J. D. Potter, N. M. Lindor, J. L. Hopper, M. A. Jenkins, A. K. Win, "Type 2 diabetes mellitus, blood cholesterol, triglyceride and

colorectal cancer risk in Lynch syndrome," *British Journal of Cancer*, vol. 121, pp. 869–876, 2019.

[12] B. Trabert, C. A. Hathaway, M. S. Rice, E. B. Rimm, P. M. Sluss, K. L. Terry, O. A. Zeleznik, S. S. Tworoger, "Ovarian cancer risk in relation to blood cholesterol and triglycerides," *Cancer Epidemiology, Biomarkers Prevention*, vol. 30, no. 11, pp. 2044–2051, 2021, doi:10.1158/1055-9965.EPI-21-0443.

[13] T. Lofterød, E. S. Mortensen, H. Nalwoga, T. Wilsgaard, H. Frydenberg, T. Risberg, A. E. Eggen, A. McTiernan, S. Aziz, E. A. Wist, A. Stensvold, J. B. Reitan, L. A. Akslen, I. Thune, "Impact of pre-diagnostic triglycerides and HDL-cholesterol on breast cancer recurrence and survival by breast cancer subtypes," *BMC Cancer*, vol. 18, article no. 654, 2018.

[14] S. A. Nabipoorashrafi, S. A. Seyedi, S. Rabizadeh, M. Ebrahimi, S. A. Ranjbar, S. K. Reyhan, A. Meysamie, M. Nakhjavani, A. Esteghamati, "The accuracy of triglyceride-glucose (TyG) index for the screening of metabolic syndrome in adults: A systematic review and meta-analysis," *Nutrition, Metabolism Cardiovascular Diseases*, vol. 32, no. 12, pp. 2677–2688, 2022, doi:10.1016/j.numecd.2022.07.024.

[15] H.-Y. Yeh, M.-H. Pan, C.-J. Huang, S.-Y. Hong, H.-I. Yang, Y.-Y. Yang, C.-C. Huang, H.-C. Tsai, T.-H. Li, C.-W. Su, M.-C. Hou, "Triglyceride-glucose index and cancer risk: a prospective cohort study in Taiwan," *Diabetology Metabolic Syndrome*, vol. 17, p. 283, 2025, doi:10.1186/s13098-025-01531-0.

[16] H.-C. Wu, L.-C. Lee, and W.-J. Wang, "Plasmapheresis for hypertriglyceridemia: The association between blood viscosity and triglyceride clearance rate," *Journal of Clinical Laboratory Analysis*, vol. 32, no. 8, 2018, doi:10.1002/jcla.22688.

[17] J. W. Han, P. S. Sung, J. W. Jang, J. Y. Choi, and S. K. Yoon, "Whole blood viscosity is associated with extrahepatic metastases and survival in patients with hepatocellular carcinoma," *PLoS ONE*, vol. 16, no. 12, p. e0260311, 2021.

[18] M. Carlisi, R. Lo Presti, S. Mancuso, S. Siragusa, and G. Caimi, "Calculated whole blood viscosity and albumin/fibrinogen ratio in patients with a new diagnosis of multiple myeloma: Relationships with some prognostic predictors," *Biomedicines*, vol. 11, no. 3, p. 964, 2023.

[19] M. Carlisi, R. Lo Presti, S. Mancuso, S. Siragusa, and G. Caimi, "Thrombotic risk and calculated whole blood viscosity in a cohort of patients with new diagnosis of multiple myeloma," *Clinical and Applied Thrombosis/Hemostasis*, vol. 29, p. 10760296231222477, 2023, doi:10.1177/10760296231222477.

[20] M. Thapa, J. Courville, R. Leonhard, P. Buchhanolla, M. A. Sheikh, R. Shah, P. Rai, H. Chokhawala, M. I. Hossain, M. A. N. Bhuiyan, J. D. Jordan, R. E. Kelley, "The correlation of whole blood viscosity and outcome in mechanical thrombectomy for acute ischemic stroke," *Frontiers in Stroke*, , vol. 4, 2025, doi:10.3389/fstro.2025.1517343.

[21] E. I. Cekirdekci and B. Bugan, "Whole blood viscosity in microvascular angina and coronary artery disease: Significance and utility/Viscosidade sanguínea na angina microvascular e na doença arterial coronária: significado e viabilidade," *Revista Portuguesa de Cardiologia*, vol. 39, no. 1, pp. 17–23, 2020, doi:10.1016/j.repc.2019.04.008.

[22] G. Pop, D. Duncker, M. Gardien, P. Vranckx, S. Versluis, and D. Hasan, "The clinical significance of whole blood viscosity in (cardio)vascular medicine," *Clinical Hemorheology and Microcirculation*, vol. 39, no. 1–4, pp. 1–10, 2008.

[23] O. K. Baskurt and H. J. Meiselman, "Blood rheology and hemodynamics," *Seminars in Thrombosis and Hemostasis*, vol. 50, no. 6, pp. 902–915, 2024, doi:10.1055/s-0043-1777802.

[24] Y.-I. Cho and D. J. Cho, "Hemorheology and microvascular disorders," *Korean Circulation Journal*, vol. 41, no. 6, pp. 287–295, 2011, doi:10.4070/kcj.2011.41.6.287.

[25] G. D. O. Lowe, A. J. Lee, A. Rumley, J. F. Price, and F. G. R. Fowkes, "Blood viscosity and risk of cardiovascular events: The Edinburgh Artery Study," *British Journal of Haematology*, vol. 96, no. 1, pp. 168–173, Jan. 1997, doi:10.1046/j.1365-2141.1997.8532481.x.

[26] L. J. Tamariz, J. H. Young, J. S. Pankow, H.-C. Yeh, M. I. Schmidt, B. Astor, and F. L. Brancati, "Blood viscosity and hematocrit as risk factors for type 2 diabetes mellitus: The Atherosclerosis Risk in Communities (ARIC) study," *American Journal of Epidemiology*, vol. 168, no. 10, pp. 1153–1160, Nov. 2008, doi:10.1093/aje/kwn243.

[27] A. J. Lee, P. I. Mowbray, G. D. O. Lowe, A. Rumley, F. G. R. Fowkes, and P. L. Allan, "Blood viscosity and elevated carotid intima-media thickness in men and women: The Edinburgh Artery Study," *Circulation*, vol. 97, no. 15, pp. 1467–1473, Apr. 1998, doi:10.1161/01.CIR.97.15.1467.

[28] Y. Li, X. Tian, T. Liu, and R. Wang, "Association between whole blood viscosity and arterial stiffness in patients with type 2 diabetes mellitus," *Endocrine*, vol. 49, pp. 148–154, 2015, doi:10.1007/s12020-014-0394-9.

[29] G. Caimi, C. Urso, S. Brucculeri, R. Lo Presti, and M. Carlisi, "Calculated whole blood viscosity in non-diabetic subjects with asymptomatic carotid atherosclerosis: How insulin resistance may affect blood viscosity," *Clinical Hemorheology and Microcirculation*, vol. 88, no. 2, pp. 199–209, 2024, doi:10.3233/CH-221422.

[30] [11] E. Kucukal, Y. Man, A. Hill, S. Liu, A. Bode, R. An, J. Kadambi, J. A. Little, U. A. Gurkan, "Whole blood viscosity and red blood cell adhesion: Potential biomarkers for targeted and curative therapies in sickle cell disease," *American Journal of Hematology*, vol. 95, no. 11, pp. 1246–1256, 2020, doi:10.1002/ajh.25933.

[31]A. Atehortúa, P. Gkontra, M. Camacho, O. Diaz, M. Bulgheroni, V. Simonetti, M. Chadeau-Hyam, J. F. Felix, S. Sebert, K. Lekadir, "Cardiometabolic risk estimation using exposome data and machine learning," *International Journal of Medical Informatics*, vol. 179, 105209, 2023, doi:10.1016/j.ijmedinf.2023.105209.

[32] F. Shamout, T. Zhu, and D. A. Clifton, "Machine learning for clinical outcome prediction," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 116–126, 2021, doi:10.1109/RBME.2020.3033397.

[33] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015,

doi:10.1161/CIRCULATIONAHA.115.001593.

[34] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1359, Apr. 2019, doi:10.1056/NEJMra1814259.

[35] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001, doi:10.1016/S0933-3657(01)00077-X.

[36] M. Nagendran, Y. Chen, C. A. Lovejoy, A. C. Gordon, M. Komorowski, H. Harvey, J. P. A. Ioannidis, G. S. Collins, M. Maruthappu, "Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies," *BMJ*, vol. 368, Art. no. m689, Mar. 2020, doi:10.1136/bmj.m689.

[37] M. U. Oner, Y.-C. Cheng, H. K. Lee, and W.-K. Sung, "Training machine learning models on patient-level data segregation is crucial in practical clinical applications," *medRxiv*, preprint, Apr. 25, 2020, doi:10.1101/2020.04.23.20076406.

[38] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Communications*, vol. 15, Art. no. 1829, 2024, doi:10.1038/s41467-024-42381-8.

[39] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *arXiv preprint arXiv:2207.07048*, 2022.

[40] A. C. Filho, A. F. D. M. Batista, and H. G. dos Santos, "Data leakage in health outcomes prediction with machine learning: Comment on 'Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning'," *Journal of Medical Internet Research*, vol. 23, no. 2, Art. no. e10969, Feb. 2021, doi:10.2196/10969.

[41]L. Sasse, E. Nicolaisen-Sobesky, J. Dukart, S. B. Eickhoff, M. Götz, S. Hamdan, V. Komeyer, A. Kulkarni, J. M. Lahnakoski, B. C. Love, F. Raimondo, K. R. Patil, "Overview of leakage scenarios in supervised machine learning," *Journal of Big Data*, vol. 12, Art. no. 135, May 2025, doi:10.1186/s40537-025-01193-8.

[42] W. Zhang, S. Tople, and O. Ohrimenko, "Leakage of dataset properties in multi-party machine learning," in *Proc. 30th USENIX Security Symposium (USENIX Security '21)*, Aug. 2021, pp. 2687–2704.

[43] Z. Li and Y. Zhang, "Membership leakage in label-only exposures," in *Proc. ACM SIGSAC Conf. Computer & Communications Security (CCS '21)*, Nov. 2021, pp. 880–895, doi:10.1145/3460120.3484575.

[44] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, New York, NY, USA: ACM, 2005, pp. 625–632, doi:10.1145/1102351.1102430.

[45] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, Sydney, Australia: PMLR, vol. 70, pp. 1321–1330, 2017.

[46] J. E. Dalton, "Flexible recalibration of binary clinical prediction models," *Statistics in Medicine*, vol. 31, no. 24, pp. 2769–2786, 2012, doi:10.1002/sim.5544.

[47] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, pp. 4765–4774, 2017.

[48] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, M. W. Kattan, "Assessing the performance of prediction models: a framework for traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, Jan. 2010, doi:10.1097/EDE.0b013e3181c30fb2.

[49] A. J. Vickers, B. Van Calster, and E. W. Steyerberg, "Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests," *BMJ*, vol. 352, Art. no. i6, Jan. 2016, doi:10.1136/bmj.i6.

[50] K. G. M. Moons, D. G. Altman, J. B. Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W. Steyerberg, A. J. Vickers, D. F. Ransohoff, and G. S. Collins, "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration," *Annals of Internal Medicine*, vol. 162, no. 1, pp. W1–W73, Jan. 2015, doi:10.7326/M14-0698.

[51] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, article no. 160018, 2016. doi:10.1038/sdata.2016.18.

[52] V. Stodden, J. Seiler, and Z. Ma, "An empirical analysis of journal policy effectiveness for computational reproducibility," *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. 2584–2589, Mar. 2018, doi:10.1073/pnas.1708290115.

[53] H. Semmelrock, T. Ross-Hellauer, S. Kopeinik, D. Theiler, A. Haberl, S. Thalmann, and D. Kowald, "Reproducibility in machine-learning-based research: Overview, barriers, and drivers," *AI Magazine*, 2025, doi:10.1002/aaai.70002.

[54] O. E. Gundersen, K. Coakley, C. Kirkpatrick, and Y. Gil, "Sources of irreproducibility in machine learning," *Information and Software Technology*, vol. 149, p. 106997, 2022.

[55] A. D. Meid, A. Gerharz, and A. Groll, "Machine learning for tumor growth inhibition: Interpretable predictive models for transparency and reproducibility," *Frontiers in Pharmacology*, vol. 13, p. 831431, 2022, doi:10.3389/fphar.2022.831431.

[56] X. Huang, Q. He, H. Hu, H. Shi, X. Zhang, and Y. Xu, "Integrating machine learning and nontargeted plasma lipidomics to explore lipid characteristics of premetabolic syndrome and metabolic syndrome," *Frontiers in Endocrinology*, vol. 14, p. 1136410, 2023, doi:10.3389/fendo.2023.1136410.

[57] D. Shin, "Prediction of metabolic syndrome using machine learning approaches based on genetic and nutritional factors: A 14-year prospective-based cohort study," *BMC Medical Genomics*, vol. 17, p. 224, 2024, doi:10.1186/s12920-024-01998-1.

[58] T. Rönn, A. Perfilyev, N. Oskolkov, and C. Ling, "Predicting type 2 diabetes via machine learning integration of multiple omics from human pancreatic islets," *Diabetologia*, vol. 64, pp. 1324–1334, 2021, doi:10.1007/s00125-021-05427-0.

[59] J. Duan, Y. Wang, L. Chen, C. L. P. Chen, and R. Zhang, "Employing broad learning and non-invasive risk factors to improve the early diagnosis of metabolic syndrome," *Frontiers in Endocrinology*, vol. 14, p. 1220131, 2023, doi:10.3389/fendo.2023.1220131.

[60] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef, "Deep generative modeling for single-cell transcriptomics," *Nature Methods*, vol. 15, pp. 1053–1058, Nov. 2018, doi:10.1038/s41592-018-0229-2.

[61] G. S. Collins, J. B. Reitsma, D. G. Altman, K. G. M. Moons, "Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement," *British Journal of Surgery*, vol. 102, no. 3, pp. 148–158, 2015, doi:10.1002/bjs.9736.

[62] G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, M. Ghassemi, X. Liu, J. B. Reitsma, M. van Smeden, A.-L. Boulesteix, J. C. Camaradou, L. A. Celi, S. Denaxas, A. K. Denniston, B. Glocker, R. M. Golub, H. Harvey, G. Heinze, M. M. Hoffman, A. P. Kengne, E. Lam, N. Lee, E. W. Loder, L. Maier-Hein, B. A. Mateen, M. D. McCradden, L. Oakden-Rayner, J. Ordish, R. Parnell, S. Rose, K. Singh, L. Wynants, P. Logullo, "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, e078378, 2024, doi:10.1136/bmj-2023-078378.

[63] J. Gallifant, M. Afshar, S. Ameen, Y. Aphinyanaphongs, S. Chen, G. Cacciamani, D. Demner-Fushman, D. Dligach, R. Daneshjou, C. Fernandes, L. H. Hansen, A. Landman, L. Lehmann, L. G. McCoy, T. Miller, A. Moreno, N. Munch, D. Restrepo, G. Savova, R. Umeton, J. W. Gichoya, G. S. Collins, K. G. M. Moons, L. A. Celi, D. S. Bitterman, "The TRIPOD-LLM reporting guideline for studies using large language models," *Nature Medicine*, vol. 31, pp. 60–69, 2025. doi:10.1038/s41591-024-03425-5.

[64] X. Liu, S. Cruz Rivera, D. Moher, M. J. Calvert, A. K. Denniston, on behalf of the SPIRIT-AI and CONSORT-AI Working Group, "Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension," *The Lancet Digital Health*, vol. 2, no. 10, pp. e537–e548, 2020, doi:10.1016/j.xops.2020.100093.

[65] H. Ibrahim, X. Liu, S. Cruz Rivera, D. Moher, A.-W. Chan, M. R. Sydes, M. J. Calvert, A. K. Denniston, "Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines," *Trials*, vol. 22, article no. 11, 2021, doi:10.1186/s13063-020-04961-6.

[66] S. H. Park, C. H. Suh, "Reporting guidelines for artificial intelligence studies in healthcare (for both conventional and large language models): what's new in 2024," *Korean Journal of Radiology*, vol. 25, no. 8, pp. 687–690, 2024, doi:10.3348/kjr.2024.0598.

[67] P. Dhiman, J. Ma, C. L. Andaur Navarro, B. Speich, G. Bullock, J. A. A. Damen, L. Hooft, S. Kirtley, R. D. Riley, B. Van Calster, K. G. M. Moons, G. S. Collins, "Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review," *BMC Medical Research Methodology*, vol. 22, article no. 101, 2022, doi:10.1186/s12874-022-01624-9.

[68] G. S. Collins, M. Chester-Jones, S. Gerry, J. Ma, J. Matos, J. Sehjal, B. Tsegaye, P. Dhiman, "Clinical prediction models using machine learning in oncology: challenges and recommendations," *BMJ Oncology*, vol. 4, no. 1, e000914, 2025, doi:10.1136/bmjonc-2025-000914.

[69] G. Varoquaux, V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," *npj Digital Medicine*, vol. 5, article no. 48, 2022, doi:10.1038/s41746-022-00599-2.

[70] Y.-Q. Cai, D.-X. Gong, L.-Y. Tang, Y. Cai, H.-J. Li, T.-C. Jing, M. Gong, W. Hu, Z.-W. Zhang, X. Zhang, G.-W. Zhang, "Pitfalls in developing machine learning models for predicting cardiovascular diseases: challenge and solutions," *Journal of Medical Internet Research*, vol. 26, e47645, 2024, doi:10.2196/47645.

[71] P. Wongyikul, P. Phinyo, P. Seephueng, K. Tanasombatkul, T. Kawamatawong, C. Wongsa, T. Thongngarm, "Methodological and applicability pitfalls of clinical prediction models for asthma diagnosis: a systematic review and critical appraisal of evidence," *BMC Medical Research Methodology*, vol. 25, article no. 232, 2025 (in press).

[72] V. C. Pezoulas, D. I. Zaridis, E. Mylona, C. Androutsos, K. Apostolidis, N. S. Tachos, D. I. Fotiadis, "Synthetic data generation methods in healthcare: a review on open-source tools and methods," *Computational and Structural Biotechnology Journal*, 2024, doi:10.1016/j.csbj.2024.07.005.

[73] M. Rujas, R. Martín Gómez del Moral Herranz, G. Fico, B. Merino-Barbancho, "Synthetic data generation in healthcare: a scoping review of reviews on domains, motivations, and future applications," *International Journal of Medical Informatics*, 2024, article no. 105763, doi:10.1016/j.ijmedinf.2024.105763.

[74] J. M. Mendes, A. Barbar, M. Refaie, "Synthetic data generation: a privacy-preserving approach to accelerate rare disease research," *Frontiers in Digital Health*, vol. 7, 2025, doi:10.3389/fdgth.2025.1563991.

[75] H. Rao, W. Liu, H. Wang, I.-C. Huang, Z. He, X. Huang, "A scoping review of synthetic data generation for biomedical research and applications," *arXiv*, 2025, arXiv:2506.16594, doi:10.48550/arXiv.2506.16594.

[76] M. Anand, K. R. Rajagopal, "A short review of advances in the modelling of blood rheology and clot formation," *Fluids*, vol. 2, no. 3, article no. 35, 2017, doi:10.3390/fluids2030035.