

Table S1 Performance in repeated stratified cross-validation (5 folds $\times$ 20 repeats; 100 fits per model) on the full cohort, reported as mean $\pm$ standard deviation

Model	AUC	Accuracy	Precision	Sensitivity	Specificity	F1
Logistic regression	0.799 $\pm$ 0.040	0.720 $\pm$ 0.040	0.707 $\pm$ 0.045	0.695 $\pm$ 0.073	0.743 $\pm$ 0.055	0.698 $\pm$ 0.048
Decision tree	0.756 $\pm$ 0.053	0.716 $\pm$ 0.045	0.722 $\pm$ 0.066	0.653 $\pm$ 0.086	0.771 $\pm$ 0.076	0.681 $\pm$ 0.058
Random forest	0.822 $\pm$ 0.040	0.761 $\pm$ 0.043	0.763 $\pm$ 0.058	0.717 $\pm$ 0.076	0.800 $\pm$ 0.063	0.737 $\pm$ 0.048
XGBoost	0.828 $\pm$ 0.038	0.761 $\pm$ 0.045	0.773 $\pm$ 0.060	0.699 $\pm$ 0.083	0.815 $\pm$ 0.060	0.731 $\pm$ 0.058
LightGBM	0.807 $\pm$ 0.040	0.743 $\pm$ 0.043	0.741 $\pm$ 0.055	0.704 $\pm$ 0.078	0.779 $\pm$ 0.063	0.719 $\pm$ 0.050
SVM	0.785 $\pm$ 0.040	0.705 $\pm$ 0.043	0.692 $\pm$ 0.048	0.674 $\pm$ 0.081	0.732 $\pm$ 0.060	0.680 $\pm$ 0.053

AUC, area under the receiver operating characteristic curve; SVM, support vector machine. The discrimination of the top models (XGBoost, random forest and LightGBM) was closely clustered and statistically indistinguishable, with a fold-to-fold standard deviation of approximately 0.04.

Table S2 Multicollinearity and robustness of feature-importance ranking

Table S2A Pearson correlation matrix of the seven predictors

	Sex	BMI	WC	WHtR	UA	TyG-WC	TyG-BMI
Sex	1.00	-0.12	0.04	0.00	0.03	0.06	-0.09
BMI	-0.12	1.00	0.12	0.08	0.03	0.13	0.92
WC	0.04	0.12	1.00	0.09	0.06	0.90	0.13
WHtR	0.00	0.08	0.09	1.00	0.01	0.14	0.12
UA	0.03	0.03	0.06	0.01	1.00	0.04	0.02
TyG-WC	0.06	0.13	0.90	0.14	0.04	1.00	0.30
TyG-BMI	-0.09	0.92	0.13	0.12	0.02	0.30	1.00

UA, serum uric acid. Strong correlations (bold in text) were confined to BMI–TyG-BMI ($r = 0.92$) and WC–TyG-WC ($r = 0.90$).

Table S2B Variance inflation factors (VIF)

Predictor	VIF
Sex	1.02
BMI	155.34
WC	121.35
WHtR	1.03
Serum uric acid	1.01
TyG-WC	155.44
TyG-BMI	191.18

High VIFs were limited to the two collinear pairs (BMI/TyG-BMI and WC/TyG-WC). WHtR, serum uric acid and sex were nearly independent ($VIF \approx 1.0$).

Table S2C Feature importance from two SHAP-independent methods (final LightGBM model)

Predictor	Permutation importance (ΔAUC) , mean $\pm$ SD	Gain importance
WHtR	0.112 $\pm$ 0.026	425.2
BMI	0.060 $\pm$ 0.015	349.4
WC	0.030 $\pm$ 0.019	280.4
TyG-WC	0.028 $\pm$ 0.016	314.8
TyG-BMI	0.022 $\pm$ 0.010	165.8
Serum uric acid	0.011 $\pm$ 0.011	206.5
Sex	0.005 $\pm$ 0.014	91.0

Permutation importance is the mean decrease in test-set AUC over 100 permutations. WHtR ranked first under both methods, reproducing the SHAP-based ranking.

Table S2D Incremental discrimination of nested feature sets (repeated cross-validation)

Feature set	Mean AUC±SD	Increment
BMI only	0.678±0.053	—
BMI + WHtR	0.758±0.045	+0.079
Anthropometric (sex, BMI, WC, WHtR)	0.802±0.041	+0.044
Full 7 (+ uric acid, TyG-WC, TyG-BMI)	0.806±0.040	+0.004

Adding WHtR to BMI markedly improved discrimination (+0.079), whereas the metabolic-derived indices added only a negligible increment beyond the anthropometric set (+0.004).