Am J Transl Res 2026;18(10):9078-9094
www.ajtr.org /ISSN:1943-8141/AJTR0174962

Original Article

Construction and validation of a cross-sectional classification model for identifying adverse metabolic status in patients with polycystic ovary syndrome based on serum amino acids and lipid metabolic characteristics

Yan Zhao[1*], Haiyang Wang[2*], Cuina Zhang[3], Weiguang Li[1]

[1]Department of laboratory, The Fourth Hospital of Shijiazhuang, Shijiazhuang 050031, Hebei, China; [2]Operating Room, Hebei Province Second Rongjun Youfu Hospital, Shijiazhuang 050051, Hebei, China; [3]Department of Nephrology with Integrated Chinese and Western Medicine, Shijiazhuang Hospital of Traditional Chinese Medicine, Shijiazhuang 050051, Hebei, China. [*]Equal contributors and co-first authors.

Received June 24, 2026; Accepted September 5, 2026; Epub October 15, 2026; Published October 30, 2026

Abstract: Objectives: To develop and externally validate a cross-sectional classification model for identifying adverse metabolic status at diagnosis in patients with polycystic ovary syndrome (PCOS) based on serum amino acid and lipid metabolic profiles. Methods: This retrospective study included 442 PCOS patients, comprising a training set (n=309; June 2022-December 2024) and an external validation set (n=133; January 2025-December 2025). Within each dataset, patients were classified into an adverse metabolic status group or a metabolically normal group according to the presence of at least one of the following conditions at baseline: metabolic syndrome, insulin resistance (IR), non-alcoholic fatty liver disease, impaired glucose tolerance, or type 2 diabetes mellitus (T2DM). At enrollment, serum concentrations of branched-chain amino acids (BCAAs), aromatic amino acids (AAAs), glucogenic amino acids, lysophosphatidylcholine (LPC) species, and free fatty acids (FFAs) were measured using liquid chromatography-tandem mass spectrometry. Routine biochemical and hormonal parameters were also collected. Results: Multivariable logistic regression identified BMI ≥ 28.215 kg/m^2 (OR=11.27), HOMA-IR ≥ 3.42 (OR=4.61), TyG index ≥ 4.84 (OR=5.44), valine ≥ 270.125 nmol/mL (OR=5.68), glutamate ≥ 63.185 nmol/mL (OR=8.87) and arachidonic acid ≥ 21.535 μmol/L (OR=10.85) as independent positive predictors of adverse metabolic status. In contrast, higher SHBG (OR=0.31), BCAA/AAA ratio ≥ 3.215 (OR=0.20), LPC(16:0) ≥ 142.06 μmol/L (OR=0.22), and LPC(18:1) ≥ 30.515 μmol/L (OR=0.28) were protective factors. The nomogram achieved areas under the receiver operating characteristic curve (AUCs) of 0.965 in the training set and 0.939 in the external validation set. Conclusion: The nomogram integrating serum amino acids and lipid profiles with routine clinical measures may provide a reliable tool for classifying adverse metabolic status in PCOS patients at initial assessment.

Keywords: Polycystic ovary syndrome, unfavorable metabolic outcomes, amino acid profiles, lipid levels, risk assessment model, nomogram

Introduction

Polycystic ovary syndrome (PCOS) affected about 69.47 million women of reproductive age worldwide in 2021, and its prevalence continues to increase, especially among adolescents and young adults [1, 2]. PCOS not only interferes with reproductive functions but also significantly increases the risk of metabolic problems, including insulin resistance, dyslipidemia, metabolic syndrome, non-alcoholic fatty liver disease, and type 2 diabetes (T2DM) [3-5]. Meta-analyses have consistently shown that metabolic syndrome is more prevalent among women with PCOS than those without PCOS, with the odds potentially being several times higher [6]. Dyslipidemia, characterized by elevated triglyceride levels and reduced high-density lipoprotein cholesterol (HDL-C), is one of the most common metabolic abnormalities in this population [7, 8]. These metabolic disturbances not only exacerbate reproductive manifesta-

https://doi.org/10.62347/AXEX6570

tions but also increase the long-term risk of cardiovascular disease, making PCOS a major public health concern.

In clinical practice, there remains a substantial unmet need for the accurate and timely identification of PCOS patients with adverse metabolic status. Current risk assessment approaches rely primarily on conventional clinical and biochemical measures. However, these approaches may not fully capture diverse metabolic abnormalities or only raise alarms when organ damage has already occurred, thereby limiting opportunities for timely intervention [9, 10]. Emerging evidence suggests that changes in amino acid and lipid metabolism may occur before overt metabolic abnormalities become clinically apparent, indicating their potential utility as early metabolic markers. Recent metabolomics studies have shown higher circulating levels of branched-chain amino acids (BCAAs) and aromatic amino acids (AAAs) in women with PCOS compared with those without PCOS [11, 12]. Importantly, these alterations are strongly associated with body mass index (BMI), insulin resistance (IR) as assessed by the homeostatic model assessment of insulin resistance (HOMA-IR), and hyperandrogenemia, indicating that dysregulated amino acid metabolism may contribute to metabolic deterioration in PCOS [12].

Phospholipid metabolism has also emerged as an important dysregulated pathway in PCOS. Multi-omics studies have suggested that glycerophospholipid metabolism is among the most changed metabolic pathways in PCOS patients [6, 13]. In particular, decreased lysophosphatidylcholine (LPC) levels and increased free fatty acids levels have been reported in the blood and follicular fluid of women with PCOS. These changes in lipids are closely associated with IR and ovarian dysfunction [14, 15]. Collectively, alterations in amino acid and phospholipid metabolism may reflect key biological processes underlying adverse metabolic outcomes in PCOS. However, existing prediction tools have not integrated these two classes of metabolites into a clinically applicable model.

To address this gap, we developed and externally validated a cross-sectional risk classification model incorporating serum amino acid and phospholipid profiles to identify adverse metabolic status at diagnosis in PCOS patients. The development and validation of this model may facilitate identification of high-risk patients, thereby supporting individualized clinical management and potentially reducing the long-term metabolic burden of PCOS.

Methods

Study design and case selection

This retrospective study included 442 patients diagnosed with PCOS who presented to the Fourth Hospital of Shijiazhuang between June 2022 and December 2025. Clinical data were extracted from the electronic medical record collection system. Inclusion criteria: (1) Fulfillment of the diagnostic criteria for PCOS [16]; (2) Age between 18 and 45 years; (3) No use of hormonal medications, including oral contraceptives, estrogen, or progesterone, or insulin-sensitizing agents, such as metformin or thiazolidinediones, within 3 months prior to treatment; (4) Availability of complete clinical and laboratory data. Exclusion criteria: (1) Other endocrine disorders known to affect reproductive or metabolic health, including Cushing's syndrome, congenital adrenal hyperplasia, androge-secreting tumors, thyroid disorders, hyperprolactinaemia, or type 1 or type 2 diabetes mellitus diagnosed before PCOS diagnosis; (2) Severe hepatic or renal dysfunction; (3) Cardiovascular diseases, such as coronary heart disease or heart failure; (4) Acute or chronic inflammatory conditions, autoimmune diseases, or active infections within 4 weeks before study enrolment; (5) Pregnancy or breastfeeding; or (6) A history of ovarian surgery, chemotherapy, or pelvic radiotherapy.

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Institutional Review Board/Ethics Committee of the Fourth Hospital of Shijiazhuang. The requirement for informed consent was waived because the study employed anonymized data obtained from existing medical records and was considered to pose minimal risk to participants, without affecting their clinical management or care. All patient data were de-identified before analysis to protect patient confidentiality and privacy.

Grouping criteria

This study employed a temporal validation design. The model developed in this study was intended as a cross-sectional classification tool for identifying patients with concurrent adverse metabolic status at enrollment. According to the date of hospital admission, eligible patients were divided into two cohorts. Patients admitted between June 2022 and December 2024 was assigned to the training cohort for model development and internal validation, whereas patients admitted between January 2025 and December 2025 constituted the temporal validation cohort for evaluating the final model. Within each cohort, patients were further classified into an adverse metabolic status group or a metabolically normal group according to the presence or absence of predefined metabolic abnormalities at enrollment. The training cohort comprised 309 patients, including 135 with adverse metabolic status and 174 with normal metabolic status. The validation cohort comprised 133 patients, including 60 with adverse metabolic status and 73 with normal metabolic status.

Adverse metabolic status was defined as the presence of at least one of the following conditions at enrollment [17]: metabolic syndrome (MetS), insulin resistance (IR), non-alcoholic fatty liver disease (NAFLD), impaired glucose tolerance (IGT) or T2DM. Although these conditions differ in their pathophysiology and clinical severity, a composite definition was adopted to reflect the frequent coexistence of metabolic disturbances in PCOS patients and to develop a clinically practical tool for identifying patients with an elevated metabolic burden. However, this composite definition may introduce outcome heterogeneity; therefore, the model should be considered a general screening tool rather than a phenotype-specific classifier.

The following internationally recognized diagnostic criteria were used to define each component of the adverse metabolic status: (1) MetS was diagnosed according to the harmonized criteria issued by the International Diabetes Federation and the American Heart Association/National Heart, Lung, and Blood Institute [18]. The presence of at least three of the following components was required: waist circumference ≥ 88 cm; fasting triglycerides ≥ 1.7 mmol/L, or use of medication for elevated triglycerides; high-density lipoprotein cholesterol (HDL-C) < 1.3 mmol/L, or treatment for reduced HDL-C levels; systolic blood pressure ≥ 130 mmHg, diastolic blood pressure ≥ 85 mmHg, or antihypertensive treatment; and fasting plasma glucose ≥ 5.6 mmol/L or previously diagnosed T2DM. (2) IR was defined as a HOMA-IR value ≥ 2.69, based on a threshold previously validated in Chinese reproductive-aged women [18]. (3) NAFLD was diagnosed based on hepatic steatosis as indicated by abdominal ultrasonography, after the exclusion of significant alcohol consumption and other causes of chronic liver disease [19]. (4) IGT was defined as a 2-h plasma glucose concentration of 7.8 to < 11.0 mmol/L during a 75-g oral glucose tolerance test, consistent with the World Health Organization criteria [19]. (5) According to the American Diabetes Association criteria, T2DM was diagnosed based on any of the following: fasting plasma glucose ≥ 7.0 mmol/L; a 2-h plasma glucose ≥ 11.1 mmol/L during a 75-g oral glucose tolerance test; glycated hemoglobin (HbA1c) ≥ 6.5% (48 mmol/mol); or random plasma glucose ≥ 11.1 mmol/L in the presence of classic symptoms of hyperglycemia [18].

Data collection and laboratory measurements

All participants underwent a standardized evaluation procedure comprising physical examination and the collection of demographic and anthropometric data. After an overnight fast of at least 8 hours, venous blood samples were collected from each participant. The samples were centrifuged at 3,000 rpm for 10 min to separate the serum, which was then aliquoted and stored at -80°C until subsequent analysis.

Routine biochemical and hormonal assays: Routine biochemical parameters were measured using a Beckman AU5800 automated biochemical analyzer (Beckman Coulter, Inc., Brea, CA, USA) with the corresponding reagents. The measured parameters included fasting plasma glucose (FPG), triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), uric acid (UA), γ-glutamyl transferase (GGT), alanine aminotransferase (ALT) and aspartate aminotransferase (AST). Fasting insulin (FINS) was measured using a chemiluminescent immunoassay on a Roche Cobas e601 analyzer (Roche Diagnostics, Basel, Switzerland). Insulin resistance was assessed using

the homeostatic model assessment of insulin resistance (HOMA-IR), calculated as HOMA-IR = (FPG × FINS)/22.5. The TyG index was calculated as follows: ln(fasting TG × FPG/2) [20]. Serum concentrations of luteinizing hormone (LH), follicle-stimulating hormone (FSH), total testosterone, and sex hormone-binding globulin (SHBG), were measured using electrochemiluminescence immunoassays on the same Roche Cobas e601 analyzer with the corresponding reagent kits.

Serum amino acids and phospholipid profiling using LC-MS/MS: Targeted metabolomic analysis of serum amino acids and phospholipids was performed using liquid chromatography-tandem mass spectrometry (LC-MS/MS). For each sample, 100 μL of serum was mixed with 300 μL of ice-cold acetonitrile containing stable isotope-labeled internal standards, including valine-d8, leucine-d3, phenylalanine-d5, and LPC(19:0) (Sigma-Aldrich, St. Louis, MO, USA). After protein precipitation and centrifugation, the supernatant was dried under a stream of nitrogen and reconstituted in water/acetonitrile (80:20, v/v) containing 0.1% formic acid before injection into the LC-MS/MS system. The analytical platform consisted of an ExionLC AD ultra-performance liquid chromatography system coupled to a Triple Quad™ 6500+ mass spectrometer with an electrospray ionization source (both from AB Sciex Pte. Ltd., Framingham, MA, USA). Chromatographic separation was performed on a Phenomenex Kinetex C18 column (100 × 2.1 mm, 2.6 μm particle size) at 40°C. Gradient elution was performed using water and acetonitrile, both containing 0.1% formic acid (LC-MS grade; Merck KGaA, Darmstadt, Germany). The mass spectrometer was operated in multiple-reaction monitoring (MRM) mode with positive electrospray ionization. The targeted analytes included three branched-chain amino acids (valine, leucine, isoleucine), three aromatic amino acids (phenylalanine, tyrosine, tryptophan), five glucogenic amino acids (alanine, glycine, serine, glutamate, proline), seven lysophosphatidylcholine species, and four free fatty acids. Analyte concentrations were determined by comparing against calibration curves established with authentic standards (amino acid standards were obtained from Sigma-Aldrich; phospholipid and free fatty acid standards from Avanti Polar Lipids, Inc., Alabaster, AL, USA). The lower limits of quantification were 1 nmol/mL for amino acids, 0.01 μmol/L for LPCs, and 0.1 μmol/L for FFAs. Intra-assay and inter-assay precision, assessed using quality-control samples, showed coefficients of variation of ≤ 10% and ≤ 15%, respectively. Data acquisition and processing was performed using Analyst 1.7.1 software (AB Sciex).

Statistical analysis and model construction

Statistical analyses were performed using SPSS version 29.0 (IBM Corp., Armonk, NY, USA) and R software version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). Continuous variables were compared using the independent-samples t-test or Mann-Whitney U test, depending on data distribution. Categorical variables were compared using the Chi-square test or Fisher's exact test, as appropriate. A two-sided *P* value below 0.05 was considered statistically significant.

A two-stage variable selection procedure was used to identify independent predictors of adverse metabolic status. First, least absolute shrinkage and selection operator (LASSO) logistic regression was performed using the *glmnet* package in R to reduce the number of candidate variables. The optimal penalty parameter (λ) was selected using 10-fold cross-validation based on the minimum binomial deviance, and only variables with non-zero coefficients were retained as potential predictors. Second, the retained variables were entered into a forward stepwise multivariable logistic regression model to identify the final independent predictors. The discriminatory ability of each predictor was assessed using the area under the receiver operating characteristic curve (AUC). A nomogram was subsequently constructed based on the final multivariable logistic model using the *rms* package in R. Model performance was evaluated in terms of discrimination using the AUC, calibration using calibration plots and the Hosmer-Lemeshow test, and clinical utility using decision curve analysis (DCA) and clinical impact curves.

To determine whether metabolic biomarkers provided incremental predictive value beyond conventional clinical indicators, two separate models were constructed. The clinical-only model included BMI, HOMA-IR, TyG index, and SHBG, whereas the combined model comprised these clinical variables together

Table 1. Baseline demographic, anthropometric, and biochemical characteristics of the patients in the training cohort

Characteristic	Normal Metabolic Status (n=174)	Adverse Metabolic Status (n=135)	t/χ^2	*P* value
Age (years)	28.43 ± 4.82	29.12 ± 5.19	1.197	0.232
BMI (kg/m^2)	26.37 ± 3.45	29.84 ± 4.12	7.879	< 0.001
Waist Circumference (cm)	84.62 ± 9.18	93.57 ± 10.46	8.000	< 0.001
Hip Circumference (cm)	96.28 ± 6.93	100.33 ± 7.58	4.897	< 0.001
SBP (mmHg)	118.34 ± 10.82	126.51 ± 12.47	6.160	< 0.001
DBP (mmHg)	76.25 ± 8.16	81.69 ± 9.53	5.401	< 0.001
HOMA-IR	2.56 ± 1.25	4.22 ± 1.98	8.522	< 0.001
TyG Index	4.57 ± 0.39	4.96 ± 0.48	7.707	< 0.001
TG (mmol/L)	1.42 ± 0.69	2.08 ± 0.94	6.862	< 0.001
TC (mmol/L)	4.68 ± 0.87	5.13 ± 0.96	4.360	< 0.001
HDL-C (mmol/L)	1.31 ± 0.28	1.06 ± 0.23	8.549	< 0.001
LDL-C (mmol/L)	2.84 ± 0.71	3.29 ± 0.86	4.842	< 0.001
UA (μmol/L)	326.47 ± 71.25	368.82 ± 78.53	4.955	< 0.001
LH (IU/L)	9.47 ± 5.23	11.03 ± 5.96	2.442	0.015
FSH (IU/L)	5.62 ± 1.44	5.31 ± 1.38	1.898	0.059
Total Testosterone (nmol/L)	1.82 ± 0.67	2.27 ± 0.89	4.889	< 0.001
SHBG (nmol/L)	36.25 ± 12.48	29.76 ± 10.35	4.995	< 0.001
Adverse metabolic outcome [n (%)]				
MetS	\	79 (58.52%)		
Insulin resistance	\	118 (87.41%)		
NAFLD	\	91 (67.41%)		
IGT	\	34 (25.19%)		
T2DM	\	19 (14.07%)		

Notes: BMI: Body Mass Index; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; HOMA-IR: Homeostatic Model Assessment for Insulin Resistance; TyG: Triglyceride-Glucose Index; TG: Triglycerides; TC: Total Cholesterol; HDL-C: High-Density Lipoprotein Cholesterol; LDL-C: Low-Density Lipoprotein Cholesterol; UA: Uric Acid; LH: Luteinizing Hormone; FSH: Follicle-Stimulating Hormone; SHBG: Sex Hormone-Binding Globulin; MetS: Metabolic syndrome; NAFLD: Non-alcoholic Fatty Liver Disease; IGT: Impaired Glucose Tolerance; T2DM: Type 2 Diabetes Mellitus.

with valine, the BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid. The AUCs of the two models were compared using DeLong's test. In addition, the categorical net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were also calculated to quantify the improvement in risk classification achieved by adding metabolic biomarkers to the clinical model. The *predictABEL* package in R was used to calculate the NRI and IDI, and their 95% confidence intervals (CIs) were estimated using bootstrap resampling with 1,000 iterations. The combined model was subsequently evaluated in the temporal validation cohort.

Sample size adequacy was assessed using the events-per-variable (EPV) criterion. The training cohort included 135 patients with adverse metabolic status, and 11 candidate predictors were retained after LASSO regression, yielding an EPV ratio of 12.3, which exceeded the recommended minimum of 10 events per predictor variable [21, 22].

Results

Baseline demographic, anthropometric, biochemical, and hormonal characteristics of the training cohort according to metabolic outcome status

Baseline characteristics of the training cohort are presented in **Table 1**. Compared with normal metabolic outcome status, patients with adverse metabolic status (n=135) had significantly higher BMI (29.84 ± 4.12 vs. 26.37 ± 3.45 kg/m^2), waist and hip circumferences,

blood pressure, HOMA-IR (4.22 ± 1.98 vs. 2.56 ± 1.25), TyG index, triglycerides, total cholesterol, LDL-C, uric acid, LH, and total testosterone levels, whereas HDL-C and SHBG were substantially lower (all $P < 0.05$). No significant between-group difference was observed in FSH levels (P=0.059).

Comparison of serum amino acid profiles between PCOS patients with adverse and normal metabolic status in the training cohort

As shown in **Figure 1**, serum amino acid profiles differed significantly between the two groups. All three branched-chain amino acids (valine, leucine, isoleucine) in the adverse metabolic group were significantly higher than the normal group (all $P < 0.001$). Among these amino acids, valine showed the largest absolute between-group difference (256.33 ± 52.19 vs. 228.47 ± 45.82 nmol/mL). Similarly, the concentrations of all three aromatic amino acids (phenylalanine, tyrosine, and tryptophan) and all five glucogenic amino acids (alanine, glycine, serine, glutamate, proline) were significantly higher in the adverse metabolic status group ($P \leq 0.013$ for serine, and $P < 0.001$ for the remaining amino acids). By contrast, the BCAA/AAA ratio was significantly lower in the adverse metabolic group (2.91 ± 0.52 vs. 3.17 ± 0.58, $P < 0.001$), suggesting a proportionally greater increase in aromatic amino acids compared with branched-chain amino acids. Overall, adverse metabolic status was associated with broadly elevated serum amino acid concentrations and a reduced BCAA/AAA ratio.

Comparison of serum LPC and free fatty acid levels between the two groups in the training cohort

As shown in **Table 2**, all seven measured LPC species (LPC(14:0), LPC(16:0), LPC(16:1), LPC(18:0), LPC(18:1), LPC(18:2) and LPC(20:4)) were significantly lower in patients with adverse metabolic outcomes. The largest absolute between-group difference was observed for LPC(16:0) (139.28 ± 29.94 vs. 157.62 ± 31.45 μmol/L, $P < 0.001$). In contrast, the concentrations of all measured free fatty acids were significantly higher in the adverse metabolic status group, including palmitic acid (214.53 ± 48.26 vs. 185.64 ± 41.87 μmol/L), oleic acid (268.35 ± 59.74 vs. 236.47 ± 52.38 μmol/L), linoleic acid (341.71 ± 74.93 vs. 308.62 ± 68.25 μmol/L), and arachidonic acid (22.16 ± 5.83 vs. 17.83 ± 4.74 μmol/L) (all $P < 0.001$). Thus, adverse metabolic status in patients with PCOS was associated with lower circulating LPC concentrations and higher free fatty acid concentrations.

Spearman correlations among clinical, amino acid, and phospholipid variables

The Spearman correlation matrix (**Figure 2**) shows the correlations between each variable and adverse metabolic status. HOMA-IR (rho= 0.435), BMI (rho=0.426), waist circumference (rho=0.410) showed the strongest positive correlations with adverse metabolic status, whereas HDL-C showed the strongest negative correlation (rho=-0.441). Among the metabolic biomarkers, glutamate (rho=0.374) and arachidonic acid (FFA 20:4, rho=0.366) showed the strongest positive correlations. Among the LPC species, LPC(16:0) (rho=-0.277) and LPC(18:1) (rho=-0.223) showed the strongest negative correlations. All these correlations were statistically significant ($P < 0.05$).

Identification of independent predictors using LASSO and multivariable logistic regression

LASSO regression was performed to select potential predictors from a large number of candidate variables (**Figure 3**). As shown in **Figure 3A**, increasing the penalty parameter λ resulted in coefficient shrinkage, with an increasing number of coefficients reduced to zero. The optimal λ was selected using 10-fold cross-validation based on the minimum binomial deviance, as indicated by the left dashed line in **Figure 3B**. Eleven candidate predictors (non-zero coefficients) were retained, including BMI, HOMA-IR, TyG index, total testosterone, SHBG, valine, BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid (FFA 20:4). These 11 variables were then entered as continuous measures into the multivariable logistic regression model to construct the final prediction model. To facilitate clinical interpretation, ROC-derived cutoff values for each predictor were subsequently used to present risk-stratification thresholds (**Table 3**). However, continuous forms of the predictors were retained in the nomogram to preserve statistical information and predictive accuracy.

Table 3 summarizes the independent predictors identified by multivariable logistic regres-

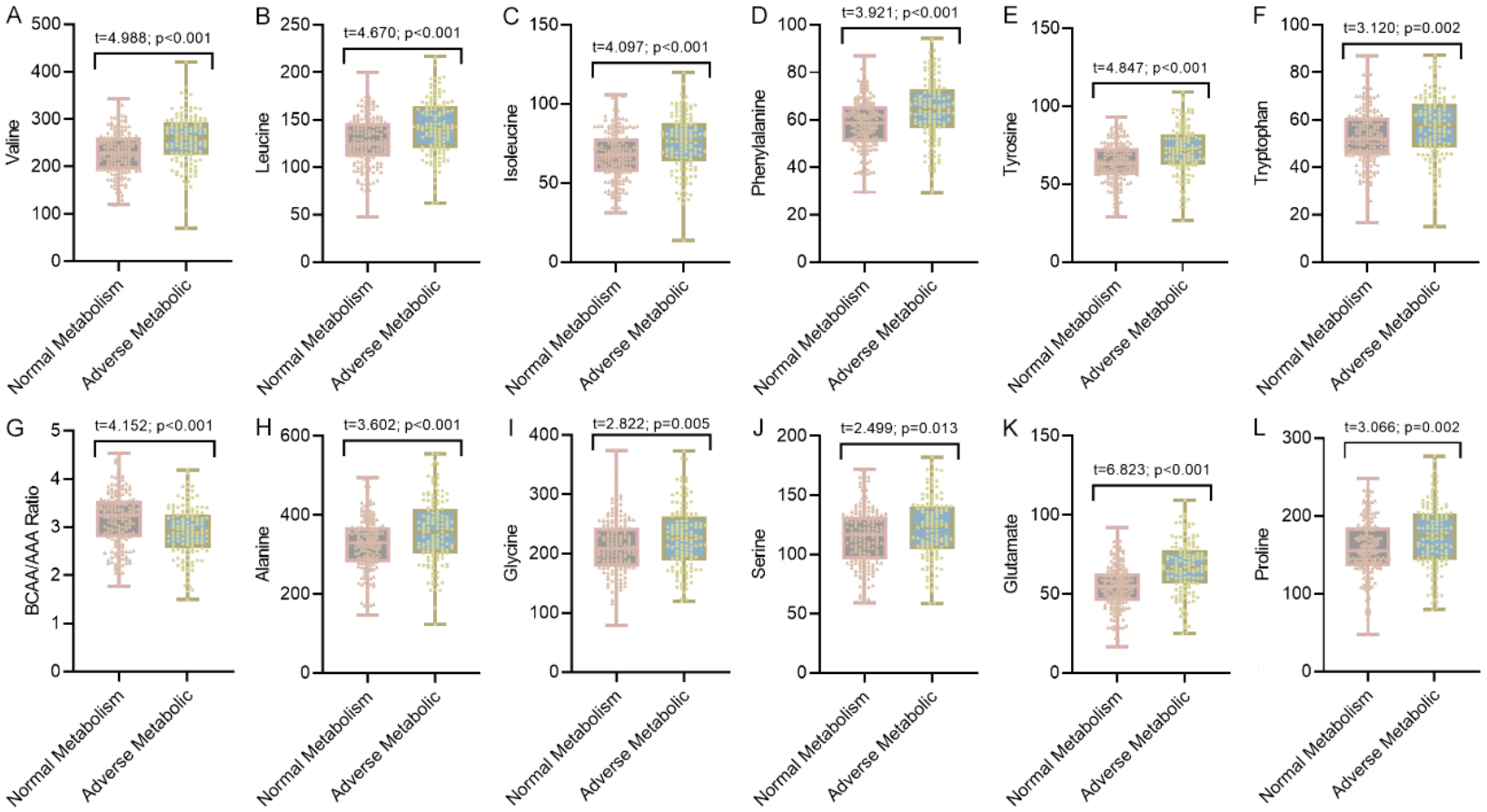

Figure 1. Comparison of serum amino acid profiles between the two cohorts in the training cohort (nmol/mL). Note: BCAA: Branched-Chain Amino Acid; AAA: Aromatic Amino Acid.

Table 2. Comparison of serum lipid profiles between the two groups in the training cohort (μmol/L)

Phospholipid Species	Normal Metabolism (n=174)	Adverse Metabolic (n=135)	*t* value	*P* value
LPCs				
LPC(14:0)	0.28 ± 0.08	0.25 ± 0.07	3.796	< 0.001
LPC(16:0)	157.62 ± 31.45	139.28 ± 29.94	5.191	< 0.001
LPC(16:1)	4.83 ± 1.52	4.27 ± 1.38	3.337	< 0.001
LPC(18:0)	48.53 ± 10.25	44.31 ± 9.86	3.651	< 0.001
LPC(18:1)	32.54 ± 8.96	28.65 ± 8.17	3.931	< 0.001
LPC(18:2)	68.29 ± 15.71	63.57 ± 14.82	2.687	0.008
LPC(20:4)	3.25 ± 1.08	2.91 ± 0.96	2.850	0.005
FFAs				
FFA (16:0) Palmitic acid	185.64 ± 41.87	214.53 ± 48.26	5.625	< 0.001
FFA (18:1) Oleic acid	236.47 ± 52.38	268.35 ± 59.74	4.989	< 0.001
FFA (18:2) Linoleic acid	308.62 ± 68.25	341.71 ± 74.93	4.049	< 0.001
FFA (20:4) Arachidonic acid	17.83 ± 4.74	22.16 ± 5.83	7.018	< 0.001

Note: LPC: Lysophosphatidylcholine; FFA: Free Fatty Acid.

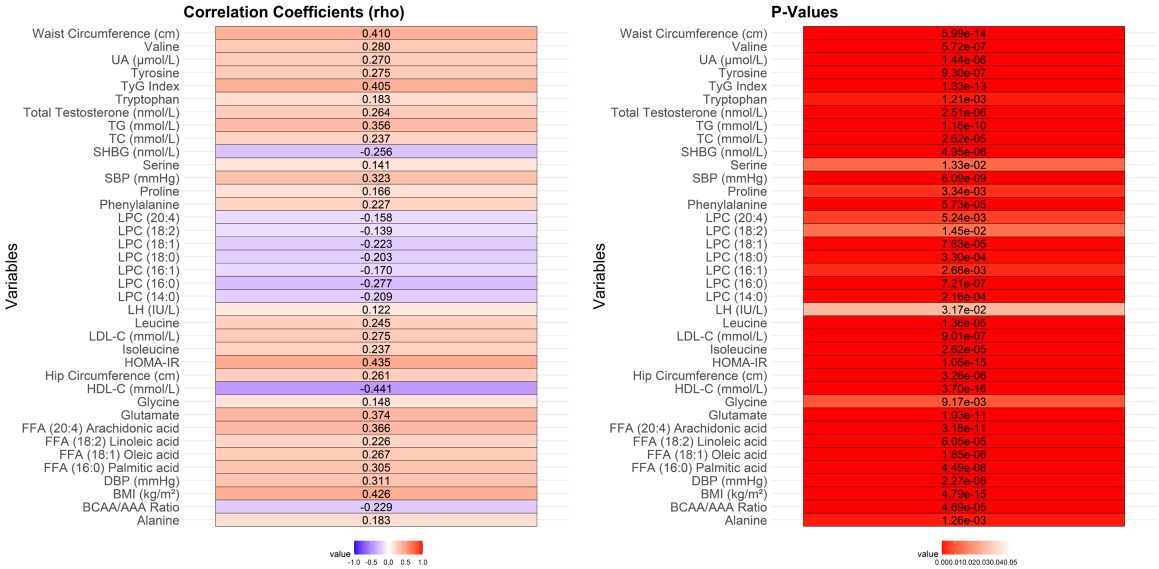

Figure 2. Spearman correlation heatmap of adverse metabolic status and measured clinical, amino acid, and lipid variables in the training cohort.

sion. After adjustment, six variables remained positively associated with adverse metabolic status: BMI ≥ 28.215 (OR=11.27, 95% CI: 4.43-28.68), HOMA-IR ≥ 3.42 (OR=4.61, 95% CI: 1.96-10.86), TyG index ≥ 4.84 (OR=5.44, 95% CI: 2.28-13.02), valine ≥ 270.125 (OR=5.68, 95% CI: 2.21-14.61), glutamate ≥ 63.185 (OR=8.87, 95% CI: 3.37-23.39), and arachidonic acid ≥ 21.535 (OR=10.85, 95% CI: 4.13-28.48). Conversely, higher SHBG levels ≥ 36.975 (OR=0.31, 95% CI: 0.12-0.78), a BCAA/AAA ratio ≥ 3.215 (OR=0.20, 95% CI: 0.08-0.49), LPC(16:0) ≥ 142.06 μmol/L (OR=0.22, 95% CI: 0.09-0.55), and LPC ≥ 30.515 μmol/L (18:1) (OR=0.28, 95% CI: 0.12-0.65) were inversely associated with adverse metabolic status (all P < 0.05).

Interaction analysis between clinical and metabolic predictors

To investigate potential interactions between clinical parameters (BMI, HOMA-IR, and TyG index) and metabolic biomarkers, multiplicative interaction terms were tested in separate logistic regression models adjusted for age and

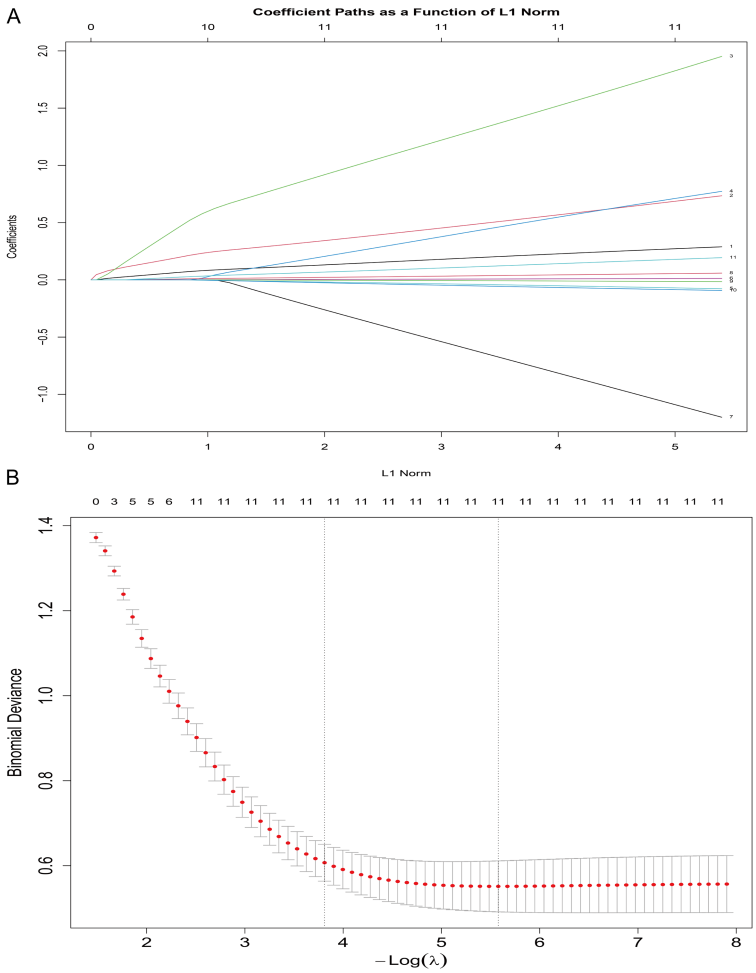

Figure 3. Feature selection process using LASSO regression. A. The characteristic changes of variable coefficients; B. The process of selecting the optimal value of parameter lambda in the Lasso regression model using the cross-validation method. Note: Lasso, least absolute shrinkage and selection operator.

other clinical covariates. None of the tested interaction terms remained statistically significant after false discovery rate (FDR) correction for multiple comparisons (all adjusted P > 0.05). Specifically, adjusted *P* values were 0.187 for the interaction between BMI and valine, 0.213 for BMI × glutamate, 0.156 for HOMA-IR × arachidonic acid, and 0.242 for TyG index × LPC(16:0). All other tested interaction terms had adjusted *P* values > 0.300 (detailed results are presented in Supplementary Table 1). These findings suggest that the predictive effects of metabolic biomarkers on adverse metabolic outcomes are largely additive rather than multiplicative in our cohort, and the selected predictors function as independent contributors to risk stratification without significant synergistic enhancement.

Establishment and evaluation of the combined model for classifying adverse metabolic status in PCOS patients

ROC analysis was used to evaluate the ability of each independent predictor to distinguish between the two adverse and normal metabolic status (**Table 4**). HOMA-IR had the highest individual AUC (0.753), followed by BMI (0.748) and the TyG index (0.735). Among the metabolic markers, glutamate (AUC=0.718) and arachidonic acid (AUC=0.713) showed moderate discriminatory ability. Valine (AUC=0.663), LPC(16:0) (AUC=0.661), LPC(18:1) (AUC=0.630), and the BCAA/AAA ratio (AUC=0.634) showed comparatively weaker discrimination.

A nomogram integrating these factors was subsequently constructed and tested (**Figure 4**). The calibration curve (**Figure 4A**) demonstrated a good agreement between the predicted probabilities and the observed metabolic status (Hosmer-Lemeshow test P > 0.05). The nomogram (**Figure 4B**) incorporated 10 predictors entered as continuous variables: BMI, HOMA-IR, TyG index, SHBG, valine, BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid. Each patient was assigned points according to their measured values, and the total score corresponded to the estimated probability of adverse metabolic status.

In the training cohort, the combined model achieved an AUC of 0.965, with a sensitivity of 0.911 and specificity of 0.902 at the optimal risk threshold of 0.47, as determined using the Youden index (**Figure 4C**). In the external validation cohort, applying this same threshold yielded a sensitivity of 0.867 and specificity of 0.877. The threshold of 0.47 may serve as a preliminary reference value for classification; however, alternative thresholds may be selected based on the specific clinical context and

Table 3. Multivariable logistic regression analysis of independent predictors for adverse metabolic status in PCOS patients in the training cohort

Variables	Coefficient	Std. Error	Wald χ^2	OR	95% CI	*P* value
BMI (kg/m^2)	2.423	0.476	5.085	11.274	4.432-28.682	< 0.001
HOMA-IR	1.528	0.437	3.492	4.607	1.955-10858	< 0.001
TyG Index	1.694	0.445	3.808	5.443	2.275-13.018	< 0.001
Total Testosterone (nmol/L)	0.873	0.466	1.873	2.394	0.960-5.968	0.061
SHBG (nmol/L)	-1.183	0.475	-2.492	0.306	0.121-0.777	0.013
Valine	1.737	0.482	3.603	5.679	2.208-14.606	< 0.001
BCAA/AAA Ratio	-1.629	0.465	-3.499	0.196	0.079-0.488	< 0.001
Glutamate	2.183	0.495	4.414	8.872	3.366-23.385	< 0.001
LPC(16:0)	-1.517	0.470	-3.226	0.219	0.087-0.551	0.001
LPC(18:1)	-1.279	0.432	-2.964	0.278	0.119-0.648	0.003
FFA(20:4) Arachidonic acid	2.384	0.492	4.842	10.848	4.133-28.476	< 0.001

Note: PCOS, polycystic ovary syndrome; BMI, body mass index; HOMA-IR, homeostatic model assessment for insulin resistance; TyG, triglyceride-glucose index; BCAA/AAA, branched-chain amino acid to aromatic amino acid ratio; LPC, lysophosphatidylcholine; FFA(20:4), arachidonic acid; SHBG, sex hormone-binding globulin.

Table 4. Receiver operating characteristic (ROC) curve analysis of each independent predictor for discriminating adverse metabolic status in PCOS patients

Variables	Best threshold	Sensitivities	Specificities	AUC	Youden index
BMI (kg/m^2)	28.215	0.689	0.730	0.748	0.419
HOMA-IR	3.420	0.659	0.759	0.753	0.418
TyG Index	4.840	0.630	0.753	0.735	0.383
Total Testosterone (nmol/L)	2.365	0.459	0.793	0.654	0.252
SHBG (nmol/L)	36.975	0.763	0.483	0.649	0.246
Valine (nmol/mL)	270.125	0.459	0.822	0.663	0.281
BCAA/AAA Ratio	3.215	0.726	0.517	0.634	0.243
Glutamate (nmol/mL)	63.185	0.63	0.764	0.718	0.394
LPC(16:0) (μmol/L)	142.06	0.600	0.684	0.661	0.284
LPC(18:1) (μmol/L)	30.515	0.644	0.632	0.630	0.276
FFA(20:4) Arachidonic acid	21.535	0.556	0.799	0.713	0.355

Notes: BMI, body mass index; HOMA-IR, homeostatic model assessment for insulin resistance; TyG, triglyceride-glucose index; BCAA/AAA, branched-chain amino acid to aromatic amino acid ratio; LPC, lysophosphatidylcholine; FFA(20:4), arachidonic acid; SHBG, sex hormone-binding globulin.

the relative priority of sensitivity versus specificity, such as lower threshold for screening purposes or higher threshold for confirmatory triage. DCA (**Figure 4D**) and the clinical impact curve (**Figure 4E**) indicated that the nomogram provided a positive net benefit across a wide range of threshold probabilities. Overall, the combined model showed greater apparent discriminatory ability than any single clinical or metabolic marker.

Incremental predictive value of metabolic biomarkers

To assess whether the addition of serum amino acid and phospholipid biomarkers significantly improved model performance beyond conventional clinical indicators alone, the clinical-only model, comprising BMI, HOMA-IR, TyG index, and SHBG, was compared with the combined model (additionally included valine, BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid) (Supplementary Table 2). The clinical-only model yielded an AUC of 0.851 (95% CI: 0.808-0.894), whereas the combined model achieved a significantly higher AUC of 0.965 (95% CI: 0.944-0.986; DeLong's test, *P* < 0.001). The categorical NRI was 0.324 (95% CI: 0.187-0.461; P < 0.001), indicating net improvement in risk classification. The IDI was 0.147 (95% CI: 0.089-0.205; P < 0.001), dem-

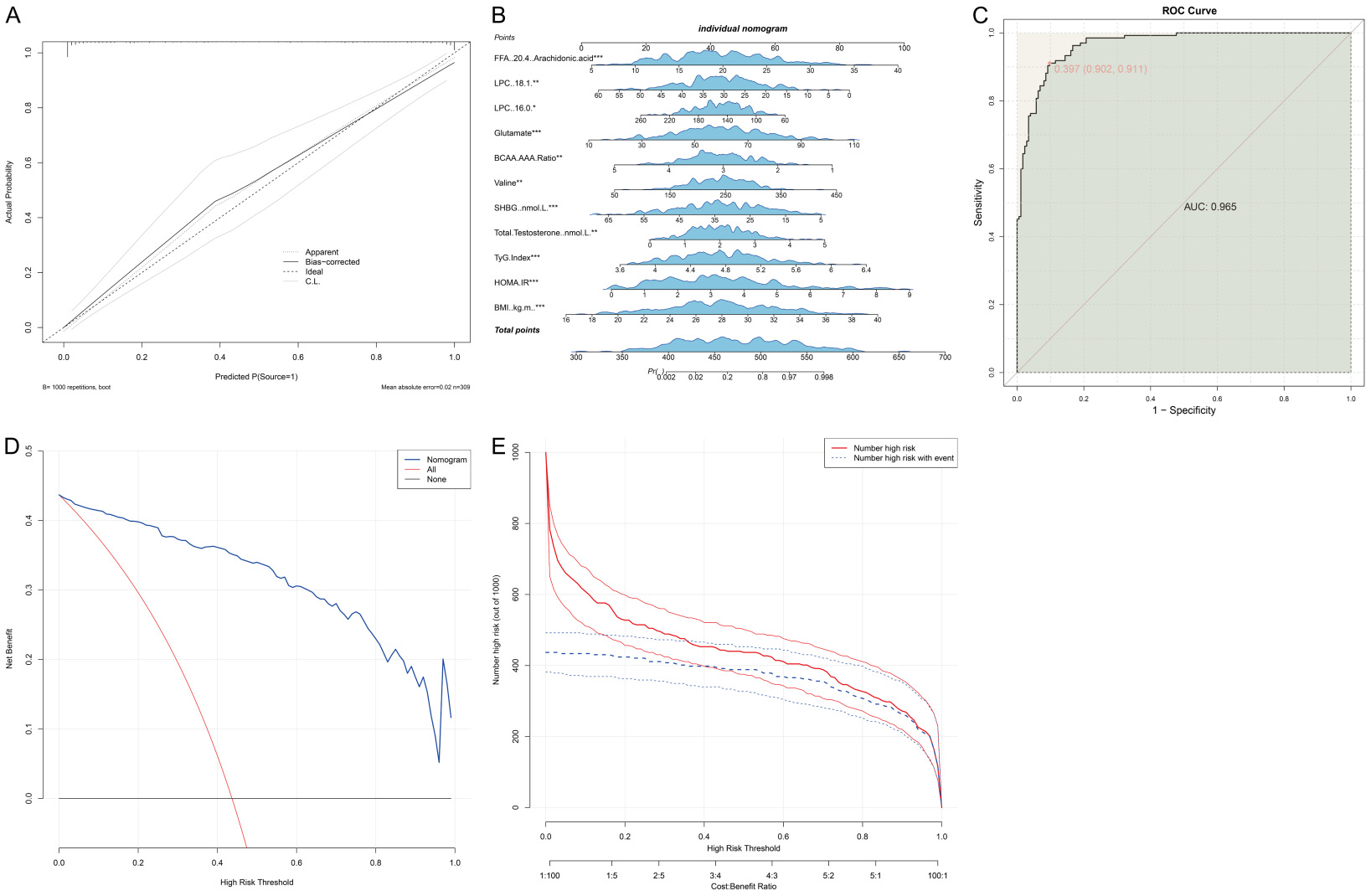

Figure 4. Development and evaluation of the integrated nomogram for classifying adverse metabolic status in PCOS patients. A. Calibration curve; B. Nomogram; C. Combined receiver operating characteristic curve (ROC); D. Decision curve analysis (DCA); E. Clinical impact curve (CIC). Note: PCOS, polycystic ovary syndrome.

Table 5. Demographic, clinical, and core predictor characteristics of the patients in the temporal validation cohort stratified by metabolic status

Characteristic	Normal Metabolic Status (n=73)	Adverse Metabolic Status (n=60)	t/χ^2	*P* value
General Data				
Age (years)	28.64 ± 4.93	29.34 ± 5.27	0.791	0.430
Waist Circumference (cm)	84.79 ± 9.24	93.82 ± 10.63	5.239	< 0.001
Hip Circumference (cm)	96.44 ± 6.98	100.42± 7.69	3.127	0.002
SBP (mmHg)	118.56 ± 10.94	126.78 ± 12.58	4.031	< 0.001
DBP (mmHg)	76.38 ± 8.21	81.92 ± 9.67	3.575	< 0.001
Adverse metabolic outcome [n (%)]				
MetS	\	35 (58.33%)		
Insulin resistance	\	52 (86.67%)		
NAFLD	\	40 (66.67%)		
IGT	\	15 (25.00%)		
T2DM	\	8 (13.33%)		
Core Predictors				
BMI (kg/m^2)	26.41 ± 3.52	28.93 ± 4.18	3.776	< 0.001
HOMA-IR	2.59 ± 1.27	4.28 ± 2.03	5.60	< 0.001
TyG Index	4.58 ± 0.40	4.97 ± 0.49	4.935	< 0.001
Total Testosterone (nmol/L)	1.84 ± 0.69	2.30 ± 0.91	3.251	0.002
SHBG (nmol/L)	36.41 ± 12.61	29.91 ± 10.47	3.187	0.002
Valine (nmol/mL)	229.12 ± 46.28	258.35 ± 53.17	3.387	< 0.001
BCAA/AAA Ratio	3.19 ± 0.59	2.93 ± 0.54	2.65	0.009
Glutamate (nmol/mL)	54.83 ± 14.02	67.21 ± 17.31	4.557	< 0.001
LPC(16:0) (μmol/L)	158.31 ± 32.17	140.12 ± 30.56	3.319	0.001
LPC(18:1) (μmol/L)	32.81 ± 9.04	28.96 ± 8.42	2.521	0.013
FFA(20:4) Arachidonic acid (μmol/L)	17.96 ± 4.81	22.43 ± 5.97	4.788	< 0.001

Notes: SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; BMI: Body Mass Index; HOMA-IR: Homeostatic Model Assessment for Insulin Resistance; TyG: Triglyceride-Glucose Index; SHBG: Sex Hormone-Binding Globulin; BCAA/AAA: Branched-Chain Amino Acid to Aromatic Amino Acid Ratio; LPC: Lysophosphatidylcholine; FFA: Free Fatty Acid.

onstrating improved separation of the predicted probabilities between patients with adverse and normal metabolic status. These findings confirm that the inclusion of the selected metabolic biomarkers provided significant incremental classification value beyond the examined conventional clinical indicators.

External validation of the combined model

The model was externally evaluated using a temporally separated cohort (**Table 5**). In the validation cohort, patients with adverse metabolic outcomes (n=60) had significantly higher waist and hip circumferences, blood pressure, BMI (28.93 ± 4.18 vs. 26.41 ± 3.52 kg/m^2), HOMA-IR (4.28 ± 2.03 vs. 2.59 ± 1.27), TyG index, total testosterone, valine, glutamate, and arachidonic acid compared with those with normal metabolic status (n=73). Conversely, SHBG, the BCAA/AAA ratio, LPC(16:0) and LPC(18:1) levels were significantly lower in the adverse metabolic status group (all $P < 0.05$). The trends of these differences were consistent with those observed in the training cohort.

Finally, the performance of the combined nomogram in the external validation cohort (n=133) was performed, as shown in **Figure 5**. The model achieved an AUC of 0.939, with specificity of 0.877 and sensitivity of 0.867, indicating good discrimination. These findings demonstrate that the model retained good discriminatory performance in a temporally distinct cohort.

Discussion

In this study, we developed and temporally validated a cross-sectional classification model

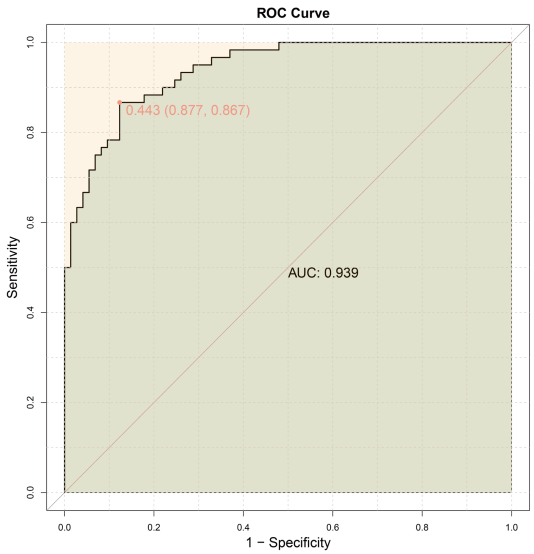

Figure 5. External validation of the nomogram in the time-external validation cohort (n=133).

that combined conventional clinical indicators with serum amino acid and lipid profiles for identifying adverse metabolic status at diagnosis in PCOS patients. The combined model, integrating anthropometric measures, surrogate markers of insulin resistance, specific amino acids, LPC species, and a free fatty acid, showed greater discriminatory ability than any single clinical or metabolic marker. The final nomogram incorporating BMI, HOMA-IR, TyG index, SHBG, valine, BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid, demonstrated good discrimination in both the training and temporal validation cohorts. This indicates that the model has robust generalizability.

Our findings are consistent with those of earlier metabolomic studies [3, 23, 24]. In our cohort, patients with adverse metabolic status exhibited broadly higher levels of branched-chain amino acids, including valine, leucine, and isoleucine; aromatic amino acids, including phenylalanine, tyrosine, tryptophan; and glucogenic amino acids, including alanine, glycine, serine, glutamate, and proline. Paczkowska et al. [11] similarly reported elevated BCAA and AAA levels in PCOS women, and these alterations were positively correlated with IR and hyperandrogenemia. Elevated circulating BCAA levels may reflect diminished BCAA catabolism associated with reduced activity of the branched-chain α-keto acid dehydrogenase complex and may contribute to overactivation of the mechanistic target of rapamycin (mTOR) signaling pathway [25, 26]. Elevated glutamate, which emerged as one of the strongest positive predictors in our model, participates in amino group transfer and metabolic pathways related to gluconeogenesis and oxidative stress; however, this interpretation is speculative because the present findings were derived from cross-sectional associations. Notably, the BCAA/AAA ratio was lower in the adverse metabolic group, reflecting a disproportionate elevation in AAAs relative to BCAAs. Similar patterns have been observed in insulin-resistant states and may be associated with altered hepatic AAA metabolism or decreased AAA clearance in the presence of hepatic steatosis [27].

Our lipidomic results showed a different pattern. Patients with adverse metabolic status had lower concentrations of all seven measured LPC species, but had higher concentrations all four FFAs. Reduced levels of LPC species, especially LPC(16:0) and LPC(18:1), have been reported in PCOS patients with IR [28-30]. Chen et al. [14] found that dysregulated LPC metabolism in granulosa cells was associated with altered estrogen receptors-mediated mTOR signaling. The observed reductions in LPC species may be associated with reduced activity of lecithin-cholesterol acyltransferase (LCAT) or increased hydrolysis by lysophospholipase D (autotaxin). Autotaxin converts LPC into lysophosphatidic acid, which is a bioactive lipid implicated in inflammatory and fibrotic processes [31]. However, the cross-sectional nature of our data cannot determine whether these LPC alterations precede or result from metabolic dysfunction. In contrast, palmitic acid, oleic acid, linoleic acid, and arachidonic acid, are known markers of fat tissue breakdown and insulin resistance [32, 33]. Through the cyclooxygenase and lipoxygenase pathways, arachidonic acid can be metabolized into pro-inflammatory eicosanoids, which have been implicated in the low-grade chronic inflammation frequently observed in PCOS [34]. Our findings demonstrated a strong association between elevated arachidonic acid levels and adverse metabolic status. Nevertheless, the cross-sectional design precludes determining whether this association reflects a causal contribution or a secondary consequence of the underlying metabolic disturbance.

Following LASSO selection and multivariable logistic regression, BMI, HOMA-IR, TyG index, valine, glutamate, and arachidonic acid were positively associated with adverse metabolic status, whereas SHBG, BCAA/AAA ratio, LPC(16:0), and LPC(18:1) were inversely associated with adverse metabolic status. The protective role of higher SHBG levels is already known, because SHBG levels are negatively related to hyperinsulinemia and free testosterone [35, 36]. The BCAA/AAA ratio remained independently associated with lower odds of adverse metabolic status after adjusting for HOMA-IR, suggesting that the relative balance between these amino acid classes may provide metabolic information beyond conventional markers. Similarly, higher LPC(16:0) and LPC(18:1) levels were associated with lower odds of adverse metabolic status. Although these findings raise the possibility that specific LPC species play an active role in maintaining insulin sensitivity and lipid homeostasis, their biological roles cannot be established from the present data. Total testosterone was positively correlated with adverse metabolic status in the univariate analysis but were no longer statistically significant in the multivariate model, possibly because its association was partly explained by other variables, such as SHBG and HOMA-IR. All findings in this study are speculative and derived from cross-sectional data. The model was developed to classify patients according to the presence of concurrent adverse metabolic status at enrollment rather than to establish temporal or causal relationships. Accordingly, the associations between amino acid and lipid metabolites and adverse metabolic status may reflect shared pathophysiological processes but do not provide evidence of causality. Longitudinal studies incorporating repeated metabolomic measurements and detailed assessments of lifestyle and environmental exposures are needed to determine whether these metabolite alterations precede the development of metabolic complications and to elucidate the directionality of the observed associations.

The principal value of our nomogram lies in its ability to facilitate the classification of metabolic status at the time of PCOS diagnosis using readily obtainable biomarkers. Conventional metabolic assessments rely on physical examination and routine laboratory tests, often overlooking patients with adverse metabolic status before overt metabolic diseases manifest. Our model incorporates amino acid and phospholipid markers, facilitating the identification of patients who warrant more comprehensive metabolic evaluation and closer clinical monitoring. For instance, a patient with elevated glutamate, valine, and arachidonic acid but reduced LPC(16:0) levels and a low BCAA/AAA ratio may warrant more further evaluation even if her HOMA-IR or TyG index is not yet significantly elevated. However, because our model uses a composite outcome, it is best suited for metabolic screening rather broad than for diagnosing any specific metabolic disorder. Clinicians should interpret the model-derived probability in conjunction with targeted assessments for individual metabolic disorders when phenotype-specific decisions are required. Decision curve analysis and clinical impact curves both indicated that the nomogram may provide a net benefit across a range of threshold probabilities, supporting further investigation of its practical application in outpatient endocrinology and reproductive medicine settings.

Several limitations of this work should be mentioned. First, this study was a retrospective, single-center study, and the model underwent temporal rather than geographic validation; therefore, site-specific biases cannot be excluded. Second, adverse metabolic status was defined using a composite endpoint comprising heterogeneous conditions, including MetS, IR, NAFLD, IGT, and T2DM. Although post-hoc sensitivity analyses suggested acceptable discrimination for the individual components, with AUCs ranging from 0.84 to 0.91, the nomogram should be interpreted as a broad screening tool rather than a condition-specific diagnostic model. Third, we only included Chinese women of reproductive age (18-45 years), limiting its generalizability to other ethnic populations or age groups. Fourth, targeted metabolomic profiling was used; untargeted metabolomic approaches may identify additional informative metabolites, such as bile acids and acylcarnitines. Fifth, the cross-sectional design precludes causal inferences, and we cannot determine whether the observed metabolic changes precede or result from the adverse outcomes. Sixth, we did not collect data on diet, physical activity, smoking, or body fat percentage, which may represent important confounders. Seventh, although no statistically significant

interactions were detected after FDR correction, modest interaction effects cannot be entirely excluded due to the limited sample size and statistical power. Eighth, despite the high AUCs (0.965 training; 0.939 validation), these estimates should be interpreted with caution. The apparent performance may partly reflect the retrospective design and residual overfitting, despite LASSO selection and bootstrap validation (optimism-corrected AUC=0.958). Furthermore, because HOMA-IR was used both as a predictor and as the criterion defining IR, which was a component of the composite outcome, incorporation bias may have contributed to overestimation of model performance. Therefore, the model should be viewed as a promising proof of concept that requires rigorous prospective, multi-center external validation.

Future prospective multicenter studies are needed to validate the utility of this model and determine whether targeting these metabolic pathways can improve long-term cardiovascular and metabolic health in women with PCOS. Specifically, longitudinal studies incorporating repeated metabolomic measurements and detailed assessment of lifestyle and environmental exposures are needed to investigate the directionality of the observed associations. Additionally, randomized controlled trials could further evaluate whether interventions that modify these metabolite pathways can reduce the incidence of adverse metabolic outcomes, thereby providing crucial evidence for causality and clinical benefit. Finally, integrating genetic and epigenetic information may improve understanding of whether these metabolic alterations contribute to or result from PCOS-related metabolic dysfunction and may help identify novel therapeutic targets.

Conclusion

In this study, we developed a classification model that integrates serum amino acid and lipid profiles with routine clinical measures to identify adverse metabolic status in PCOS patients at initial assessment. The identified metabolites, particularly glutamate, arachidonic acid, LPC(16:0), LPC(18:1) and the BCAA/AAA ratio, are associated with biological processes implicated in insulin resistance, ectopic fat accumulation, and chronic inflammation. However, given the cross-sectional design, these findings should be interpreted as associations rather than evidence of causality.

Acknowledgements

This study was supported by the Medical Science Research Project of Hebei (No. 20231686).

Disclosure of conflict of interest

None.

Address correspondence to: Yan Zhao, Department of laboratory, The Fourth Hospital of Shijiazhuang, No. 16 Tangu North Street, Chang'an District, Shijiazhuang 050031, Hebei, China. E-mail: whaiyang202304@163.com

References

[1] Ning B, Zhang M, Su N, Bur R, Ma H, Liu K and La X. Global, regional, and national burden of polycystic ovary syndrome, 1990-2021: a comprehensive analysis of trends and projections based on the global burden of disease 2021 study. Gynecol Endocrinol 2025; 41: 2535737.

[2] Zeng W, Gan D, Ou J and Tomlinson B. Global trends and health system impact on polycystic ovary syndrome: a comprehensive analysis of age-stratified females from 1990 to 2021. Front Reprod Health 2025; 7: 1642369.

[3] Arvanitakis K, Chatzikalil E, Kalopitas G, Patoulias D, Popovic DS, Metallidis S, Kotsa K, Germanidis G and Koufakis T. Metabolic dysfunction-associated steatotic liver disease and polycystic ovary syndrome: a complex interplay. J Clin Med 2024; 13: 4243.

[4] Reytor-González C, Annunziata G, Campuzano-Donoso M, Morales-López T, Basantes-Tituaña C, Fascì-Spurio F, Verde L, Muscogiuri G, Barrea L, Frias-Toral E and Simancas-Racines D. Endocrinologist's crucial role in metabolic dysfunction-associated steatotic liver disease: a comprehensive review. Minerva Endocrinol (Torino) 2025; 50: 209-226.

[5] Singh V, Mahra K, Jung D and Shin JH. Gut microbes in polycystic ovary syndrome and associated comorbidities; type 2 diabetes, non-alcoholic fatty liver disease (NAFLD), cardiovascular disease (CVD), and the potential of microbial therapeutics. Probiotics Antimicrob Proteins 2024; 16: 1744-1761.

[6] Moran LJ, Misso ML, Wild RA and Norman RJ. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. Hum Reprod Update 2010; 16: 347-363.

[7] Pourteymour Fard Tabrizi F, Alipoor B, Mehrzad Sadaghiani M, Ostadrahimi A and Malek Mahdavi A. Metabolic syndrome and its characteristics among reproductive-aged women with polycystic ovary syndrome: a cross-sectional

study in Northwest Iran. Int J Fertil Steril 2013; 6: 244-249.
[8] Zeng LH, Rana S, Hussain L, Asif M, Mehmood MH, Imran I, Younas A, Mahdy A, Al-Joufi FA and Abed SN. Polycystic ovary syndrome: a disorder of reproductive age, its pathogenesis, and a discussion on the emerging role of herbal remedies. Front Pharmacol 2022; 13: 874914.
[9] Zad Z, Jiang VS, Wolf AT, Wang T, Cheng JJ, Paschalidis IC and Mahalingaiah S. Predicting polycystic ovary syndrome with machine learning algorithms from electronic health records. Front Endocrinol (Lausanne) 2024; 15: 1298628.
[10] Guixue G, Yifu P, Yuan G, Xialei L, Fan S, Qian S, Jinjin X, Linna Z, Xiaozuo Z, Wen F and Wen Y. Progress of the application clinical prediction model in polycystic ovary syndrome. J Ovarian Res 2023; 16: 230.
[11] Paczkowska K, Rachoń D, Berg A, Rybka J, Kapczyńska K, Bolanowski M and Daroszewski J. Alteration of branched-chain and aromatic amino acid profile as a novel approach in studying polycystic ovary syndrome pathogenesis. Nutrients 2023; 15: 4153.
[12] Paczkowska K, Rachoń D, Berg A, Rybka J, Kapczyńska K, Bolanowski M and Daroszewski J. Specific alteration of branched-chain amino acid profile in polycystic ovary syndrome. Biomedicines 2023; 11: 108.
[13] Zhang F, Ding Y, Zhang B, He M, Wang Z, Lu C and Kang Y. Analysis of methylome, transcriptome, and lipid metabolites to understand the molecular abnormalities in polycystic ovary syndrome. Diabetes Metab Syndr Obes 2023; 16: 2745-2763.
[14] Chen X, Shen X, Guan S, Liu Y, Song N, Song W and Jin H. Mechanisms of c-Fos regulation of mTOR signaling via ERα/β in abnormal lipid metabolism of granulosa cells in PCOS. Front Endocrinol (Lausanne) 2025; 16: 1587595.
[15] Yan H, Wang L, Zhang G, Li N, Zhao Y, Liu J, Jiang M, Du X, Zeng Q, Xiong D, He L, Zhou Z, Luo M and Liu W. Oxidative stress and energy metabolism abnormalities in polycystic ovary syndrome: from mechanisms to therapeutic strategies. Reprod Biol Endocrinol 2024; 22: 159.
[16] Dumesic DA, Oberfield SE, Stener-Victorin E, Marshall JC, Laven JS and Legro RS. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. Endocr Rev 2015; 36: 487-525.
[17] Feng J, Luo R, Gao J, Wang Y, Cong J, Ma H and Wu X. Elevated triglyceride-glucose index is associated with insulin resistance, metabolic syndrome components, nonalcoholic fatty liver disease, and adverse pregnancy outcomes in chinese women with polycystic ovary syndrome. Clin Endocrinol (Oxf) 2025; 103: 317-326.
[18] Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM and Smith SC Jr; International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; International Association for the Study of Obesity. Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. Circulation 2009; 120: 1640-1645.
[19] De Flines J and Scheen AJ. Management of metabolic syndrome and associated cardiovascular risk factors. Acta Gastroenterol Belg 2010; 73: 261-266.
[20] Tahapary DL, Pratisthita LB, Fitri NA, Marcella C, Wafa S, Kurniawan F, Rizka A, Tarigan TJE, Harbuwono DS, Purnamasari D and Soewondo P. Challenges in the diagnosis of insulin resistance: focusing on the role of HOMA-IR and Tryglyceride/glucose index. Diabetes Metab Syndr 2022; 16: 102581.
[21] Peduzzi P, Concato J, Kemper E, Holford TR and Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49: 1373-1379.
[22] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins G and van Smeden M. Calculating the sample size required for developing a clinical prediction model. BMJ 2020; 368: m441.
[23] Prosperi S and Chiarelli F. Insulin resistance, metabolic syndrome and polycystic ovaries: an intriguing conundrum. Front Endocrinol (Lausanne) 2025; 16: 1669716.
[24] Helvaci N and Yildiz BO. Polycystic ovary syndrome as a metabolic disease. Nat Rev Endocrinol 2025; 21: 230-244.
[25] Dimou A, Tsimihodimos V and Bairaktari E. The critical role of the branched chain amino acids (BCAAs) catabolism-regulating enzymes, branched-chain aminotransferase (BCAT) and branched-chain α-Keto acid dehydrogenase (BCKD), in human pathophysiology. Int J Mol Sci 2022; 23: 4022.
[26] Szmygin H, Lenart-Lipinska M, Szydelko J, Wozniak S and Matyjaszek-Matuszek B. Branched-chain amino acids as a novel biomarker of

metabolic disturbances in women with polycystic ovary syndrome - literature review. Ginekol Pol 2022; 93: 665-669.
[27] Zhao Y, Fu L, Li R, Wang LN, Yang Y, Liu NN, Zhang CM, Wang Y, Liu P, Tu BB, Zhang X and Qiao J. Metabolic profiles characterizing different phenotypes of polycystic ovary syndrome: plasma metabolomics analysis. BMC Med 2012; 10: 153.
[28] Zhang CH, Liu XY and Wang J. Essential role of granulosa cell glucose and lipid metabolism on oocytes and the potential metabolic imbalance in polycystic ovary syndrome. Int J Mol Sci 2023; 24: 16247.
[29] Li W, Gan J, Jiao J, Li J, Xu L, Li D, Geng Y, Shen M, Liang J, Lin H and Zhong L. Gut microbiota, lipid metabolism, and PCOS: A Mendelian randomization and mediation analysis. Medicine (Baltimore) 2026; 105: e47177.
[30] Zhang M, Wang Y, Di J, Zhang X, Liu Y, Zhang Y, Li B, Qi S, Cao X, Liu L, Liu S and Xu F. High coverage of targeted lipidomics revealed lipid changes in the follicular fluid of patients with insulin-resistant polycystic ovary syndrome and a positive correlation between plasmalogens and oocyte quality. Front Endocrinol (Lausanne) 2024; 15: 1414289.
[31] Yamamoto-Mikami A, Tanaka Y, Tsutsumi T, Kuwahara A and Tokumura A. Altered ovarian tissue level of lysophosphatidic acid and mRNA expressions of its metabolic enzymes and receptors in rats received gonadotropin-hyperstimulation. Reprod Biol 2024; 24: 100849.
[32] Tian Y, Zhang J, Li M, Shang J, Bai X, Zhang H, Wang Y, Chen H and Song X. Serum fatty acid profiles associated with metabolic risk in women with polycystic ovary syndrome. Front Endocrinol (Lausanne) 2023; 14: 1077590.
[33] Escobar-Morreale HF, Martínez-García MÁ, Insenser M, Cañellas N, Correig X and Luque-Ramírez M. Serum metabolomics profiling by proton nuclear magnetic resonance spectrometry of the response to single oral macronutrient challenges in women with polycystic ovary syndrome (PCOS) compared with male and female controls. Biol Sex Differ 2023; 14: 62.
[34] Yan Y, Di F, Zhang R, Song L, Zhang R and Qiu J. Arachidonic acid inhibit granulosa cell function by affecting metabolic function of liver in brown adipose transplantation rats. J Ovarian Res 2024; 17: 43.
[35] Xing C, Zhang J, Zhao H and He B. Effect of sex hormone-binding globulin on polycystic ovary syndrome: mechanisms, manifestations, genetics, and treatment. Int J Womens Health 2022; 14: 91-105.
[36] Kochanowicz M, Rhaiem TB, Owczarek AJ, Wójtowicz M, Madej P, Chudek JT and Olszanecka-Glinianowicz M. The cut-off values for SHBG discriminating insulin resistance based on the TyG, TyG-BMI, and TyG-WC values in women with PCOS. Biomedicines 2026; 14: 187.

Supplementary Table 1. Multiplicative interaction analysis between clinical predictors and metabolic biomarkers in the multivariable logistic regression model for adverse metabolic status

Interaction Term	Adjusted OR (95% CI)	Wald χ^2	*P*-value	FDR-adjusted *P*-value
BMI × Valine	1.142 (1.021-1.281)	3.124	0.032	0.187
BMI × Glutamate	1.164 (1.023-1.324)	2.987	0.041	0.213
BMI × Arachidonic acid	1.082 (0.944-1.242)	1.454	0.182	0.421
BMI × LPC(16:0)	0.922 (0.821-1.052)	2.011	0.156	0.385
BMI × LPC(18:1)	0.932 (0.822-1.064)	0.982	0.321	0.512
BMI × BCAA/AAA ratio	0.903 (0.781-1.042)	1.892	0.169	0.405
HOMA-IR × Valine	1.214 (1.023-1.423)	4.218	0.040	0.211
HOMA-IR × Glutamate	1.182 (0.992-1.384)	3.452	0.063	0.285
HOMA-IR × Arachidonic acid	1.224 (1.052-1.423)	5.892	0.028	0.156
HOMA-IR × LPC(16:0)	0.852 (0.721-1.012)	3.121	0.078	0.312
HOMA-IR × LPC(18:1)	0.882 (0.742-1.043)	2.452	0.118	0.354
HOMA-IR × BCAA/AAA ratio	0.823 (0.681-0.992)	4.012	0.045	0.221
TyG index × Valine	1.102 (0.952-1.273)	1.891	0.169	0.405
TyG index × Glutamate	1.142 (0.972-1.342)	2.452	0.118	0.354
TyG index × Arachidonic acid	1.121 (0.954-1.312)	1.981	0.158	0.392
TyG index × LPC(16:0)	0.862 (0.742-1.012)	3.842	0.049	0.242
TyG index × LPC(18:1)	0.902 (0.782-1.053)	1.452	0.228	0.456
TyG index × BCAA/AAA ratio	0.884 (0.754-1.032)	2.652	0.104	0.342

Notes: OR, Odds Ratio; CI, Confidence Interval; FDR, False Discovery Rate (Benjamini-Hochberg method). All models were adjusted for age and other clinical covariates included in the final multivariable model. BMI, body mass index; HOMA-IR, homeostatic model assessment for insulin resistance; TyG, triglyceride-glucose index; BCAA/AAA, branched-chain amino acid to aromatic amino acid ratio; LPC, lysophosphatidylcholine. The FDR-adjusted *P*-values for all interaction terms were > 0.050, indicating no statistically significant multiplicative interactions.

Supplementary Table 2. Incremental predictive value of metabolic biomarkers

Model	Clinical-only model	Combined model
AUC (95% CI)	0.851 (0.808-0.894)	0.965 (0.944-0.986)
NRI (95% CI)	Reference	0.324 (0.187-0.461)
IDI (95% CI)	Reference	0.147 (0.089-0.205)
P value	\	< 0.001

Notes: Clinical-only model includes BMI, HOMA-IR, TyG index, and SHBG. Combined model includes clinical variables plus valine, BCAA/AAA ratio, glutamate, LPC(16:0), LPC(18:1), and arachidonic acid. AUC: Area under the curve; NRI: Net Reclassification Improvement; IDI: Integrated Discrimination Improvement.