

Identification of Core Biomarkers and Temporal Mechanism Investigation in Cardiovascular Disease: A Machine Learning and Cross-Lagged Path Model Approach

Mingdong Zhang

Xinjiang Medical University

Yapeng Huang

Xinjiang Medical University

Zhaoding Liang

Xinjiang Medical University

Pei Liu

Xinjiang Medical University

Gaofeng Sun

sgfxj2004@126.com

Department of Health Education, Urumqi Center for Disease Control and Prevention (Urumqi Municipal Health Supervision Institute)

Research Article

Keywords: MACHINE LEARNING, CARDIOVASCULAR DISEASE, TEMPORAL RELATIONSHIP, BIOMARKERS, COMORBIDITY

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10795812/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

1 Identification of Core Biomarkers and Temporal Mechanism Investigation in Cardiovascular

2 Disease: A Machine Learning and Cross-Lagged Path Model Approach

3

4 **Mingdong Zhang¹, Yapeng Huang¹, Zhaoding Liang¹, Pei Liu¹, Gaofeng Sun^{2*}**

5 ¹School of Public Health, Xinjiang Medical University, Urumqi, Xinjiang, China ;

6 ²*Department of Chronic Non-Communicable Disease Prevention and Control, Urumqi Center for*

7 *Disease Control and Prevention(Urumqi Municipal Health Supervision Institute), Urumqi,*

8 *Xinjiang, China.*

9 *Correspondence: Sun Gaofeng, E-mail:sgfxj2004@126.com

10

Abstract

Objective Utilizing the China Health and Retirement Longitudinal Study (CHARLS) database, this research aims to develop a risk prediction model for cardiovascular diseases (CVD) in Chinese residents, identify core predictive indicators, establish their causal relationships with CVD, and provide screening tools for CVD prevention and management. **Methods** National survey data from 2015 CHARLS database were analyzed. Univariate analysis first identified CVD-related factors. Subsequently, four machine learning algorithms - regularized logistic regression, random forest, support vector machine, and XGBoost - were optimized using 5-fold cross-validation. Data were partitioned into training-test sets (8:2 ratio). Optimized algorithms constructed CVD risk models while evaluating predictive performance. Participants with complete 2011 baseline data were selected for cross-lagged path analysis to establish causal relationships between core predictors and CVD. **Results** The study enrolled 10,666 participants (mean age 61.58 $\pm$ 9.43 years; 53.93% female; 63.20% rural residents) with CVD prevalence of 44.13% (4,707/10,666). Univariate analysis revealed significant differences (all $P < 0.05$) between CVD and non-CVD groups in gender, residence, education level, age, white blood cell count, mean corpuscular volume, blood urea nitrogen, fasting glucose, total cholesterol, triglycerides, HDL-C, LDL-C, C-reactive protein, HbA1c, cystatin C, TyG index, TyG-BMI index, creatinine, urea, hematocrit, and hemoglobin. The XGBoost model demonstrated optimal performance. Both cystatin C and TyG-BMI index emerged as top predictors across all models. Cross-lagged path analysis revealed bidirectional causality between CVD and TyG-BMI index, while unidirectional causality existed between CVD and cystatin C. **Conclusion** The high CVD prevalence in Chinese residents was effectively predicted by XGBoost model. TyG-BMI index and cystatin C serve as

core predictive markers. The bidirectional CVD-TyG-BMI relationship and unidirectional CVD-cystatin C association suggest comorbidity risks between cardiovascular, metabolic, and renal diseases.

Keywords: MACHINE LEARNING; CARDIOVASCULAR DISEASE; TEMPORAL RELATIONSHIP; BIOMARKERS; COMORBIDITY;

Research Insights

What is currently known about this topic?

Cardiovascular diseases (CVDs) are associated with high disability and mortality rates. Global population aging has exacerbated the disease burden of CVDs.

What is the key research question?

Which biomarkers demonstrate optimal predictive value for cardiovascular disease onset?

What is new?

This study: (1) identified high-performance predictive biomarkers for CVDs, (2) evaluated the predictive efficacy of various machine learning algorithms, and (3) investigated methods to establish causal relationships between biomarkers and disease outcomes.

How might this study influence clinical practice?

The findings may facilitate earlier detection and diagnosis of cardiovascular diseases in clinical practice, enabling timely intervention.

Introduction

Cardiovascular diseases (CVD), encompassing ischemic or hemorrhagic disorders affecting the heart, brain, and systemic tissues, primarily arise from dyslipidemia, hyperviscosity, atherosclerosis, and hypertension. These conditions are characterized by high prevalence,

disability rate, and mortality¹. By 2019, global CVD prevalence reached 523 million cases with 18.6 million deaths worldwide². The escalating CVD burden has imposed substantial socioeconomic impacts across all societal levels³. Notably, the Chinese Cardiovascular Health and Diseases Report 2022 identifies CVD as the leading cause of mortality in China, accounting for 46.74% of all deaths⁴⁵. Population aging synergizes with risk factors including diabetes mellitus, dyslipidemia, and chronic kidney disease to accelerate CVD pathogenesis⁶⁷. Current CVD risk assessment predominantly employs traditional biomarkers including lipid/glycemic metabolism indicators and inflammatory mediators⁸. Large-scale clinical trials, however, demonstrate significant limitations of single-biomarker approaches – exemplified by LDL-cholesterol (LDL-C) reduction accounting for merely 15% of adverse cardiovascular risk reduction⁹. Emerging research emphasizes composite metabolic indices and comorbidities, with increasing attention to biomarker-based predictive models. Current evidence nevertheless remains constrained by cross-sectional designs that fail to establish temporal causality between biomarkers and CVD, compounded by inadequate comparative validation across modeling frameworks.

Utilizing the China Health and Retirement Longitudinal Study (CHARLS) national cohort, this investigation integrates machine learning with cross-lagged path modeling. Through multi-algorithm optimization, we systematically evaluate biochemical indicators spanning traditional biomarkers, composite metabolic parameters, and multimorbidity markers to identify core CVD predictors and decipher their temporal causality with CVD progression.

Materials and Methods

Study Population

The data were derived from the China Health and Retirement Longitudinal Study (CHARLS)

database. Initiated in 2011 and administered by Peking University, this nationally representative survey covers 12,073 households across 450 communities/villages in 150 county-level districts spanning 28 Chinese provinces. We utilized the 2015 follow-up dataset as the primary data source for model development. After excluding participants with incomplete demographic or biochemical data, 10,666 eligible subjects were included. Target variables comprised three categories: 1) Disease status: Participants with hypertension, heart disease, or stroke were classified as CVD cases; 2) Demographics: Age, gender, residence (urban/rural), and education level; 3) Biochemical parameters: TyG index, cystatin C, HbA1c, TyG-BMI index, triglycerides, fasting glucose, C-reactive protein, uric acid, creatinine, hemoglobin, total cholesterol, hematocrit, LDL-C, white blood cell count, HDL-C, mean corpuscular volume, blood urea nitrogen, and platelet count. Following identification of core CVD predictors through modeling, participants with complete 2011 baseline data were selected for subsequent causal inference analysis.

Operational Definitions

1) Lipid thresholds (per 2016 Chinese Guidelines for Dyslipidemia Management¹⁰): Total cholesterol (TC): Optimal <200 mg/dL, Borderline elevated 200-239 mg/dL, Elevated ≥ 240 mg/dL; Triglycerides (TG): Optimal <150 mg/dL, Borderline elevated 150-199 mg/dL, Elevated ≥ 200 mg/dL; High-density lipoprotein cholesterol (HDL-C): Reduced <40 mg/dL; Low-density lipoprotein cholesterol (LDL-C): Optimal <130 mg/dL, Borderline elevated 130-159 mg/dL, Elevated ≥ 160 mg/dL.

2) Clinical reference ranges (per National Guide to Clinical Laboratory Procedures, 4th Edition¹¹): Hematocrit (HCT): Male 40-50%, Female 35-45%, Platelet (PLT) count: $125-350 \times 10^9/L$, White Blood Cell (WBC) count: $3.5-9.5 \times 10^9/L$, Cystatin C (CysC): 0.59-1.03 mg/L, Hemoglobin

A1c(HbA1c): 5-8%,Uric acid(UA): Male 3.5-7.2 mg/dL, Female 2.6-6.0 mg/dL,Creatinine(Cr): Male 0.6-1.2 mg/dL, Female 0.5-1.1 mg/dL,Hemoglobin(Hb): Male 13.0-17.5 g/dL (130-175 g/L), Female 11.5-15.0 g/dL (115-150 g/L), Mean Corpuscular Volume(MCV): 80-100 fL, Blood Urea Nitrogen(BUN): 8-23 mg/dL.

3) Diabetes diagnosis (per 2024 Chinese Diabetes Prevention Guidelines¹²):Normal fasting plasma glucose (FPG)Normal fasting glucose: <110mg/dL(<6.11mmol/L),Impaired fasting glucose(IFG): 110-125mg/dL (6.11-6.94 mmol/L),Diabetes: ≥ 126 mg/dL (≥ 7.01 mmol/L).

4) C-reactive protein(CRP): Normal range 0-8 mg/L (based on clinical consensus)

5) Composite indices : Triglyceride-glucose(TyG)index index: Median [IQR] = 8.628 [8.258, 9.097];Triglyceride-glucose-BMI(TyG-BMI) index: Median [IQR] = 205.897 [179.687, 235.361]

6) CVD classification: Participants with hypertension, coronary heart disease, or stroke were defined as CVD cases.

Statistical Analysis

Data processing and analysis were performed using SPSS (version 26.0; IBM Corp) and R (version 4.3.3; R Foundation). Chi-square tests first identified demographic and biochemical variables associated with CVD prevalence through univariate analysis. Variables demonstrating statistical significance ($P < 0.05$) subsequently underwent hyperparameter tuning via 5-fold cross-validation. The optimized variables were then incorporated into four machine learning algorithms: regularized logistic regression, random forest, support vector machine (SVM), and XGBoost, to assess their predictive contributions to CVD risk stratification. Core biochemical predictors identified through model comparison were further analyzed using cross-lagged path modeling. This longitudinal analysis leveraged complete 2011 baseline data to establish temporal

precedence and bidirectional causality between selected biomarkers and CVD outcomes.

Statistical significance was defined as two-tailed $P < 0.05$ throughout the analysis.

Results

Baseline Characteristics

The study cohort comprised 10,666 participants with a mean age of 61.58 ± 9.43 years.

Demographic distribution showed female predominance (5,752 cases, 53.93%) and higher representation of rural residents (63.20%). The majority attained primary education or lower (45.20%). Overall CVD prevalence was 44.13% (4,707/10,666).

Using CVD status as the outcome variable, univariate analysis incorporating demographic variables (gender, residence, education level, age) revealed statistically significant differences in CVD prevalence across these categories (all $P < 0.05$). Significant biochemical disparities ($P < 0.05$) between CVD and non-CVD groups were observed in: WBC count, MCV, BUN, FPG, TC, TG, HDL-C, LDL-C, CRP, HbA1c, CysC, TyG index, TyG-BMI index, Cr, UA, HCT, and Hb. PLT count showed no statistical significance ($P = 0.82$).

Table 1. Univariate Analysis of Cardiovascular Disease Risk Factors

Variable	Category	Non-CVD (n=5,959)	CVD (n=4,707)	χ^2	P -value
Gender	Female	3,105	2,647	18.05	<0.001
	Male	2,854	2,060		
Residence	Urban	2,105	1,820	12.62	<0.001
	Rural	3,854	2,887		
Education	≤Primary	2593	2228	26.218	<0.001
	Primary	1359	1073		
	Middle school	1369	903		
	≥High school	638	503		
Age	<45 岁	108	43	244.981	<0.001
	45~65 岁	4256	2738		
	>65 岁	1595	1926		
WBC	<3.5×/L	246	165	16.178	<0.001
	3.5-9.5×/L	5532	4335		

	>9.5×/L	181	207		
	<80fL	391	243	16.577	<0.001
MCV	80-100fL	5023	4097		
	>100fL	545	367		
	<125×/L	719	533	1.539	0.463
PLT	125-350×/L	5103	4070		
	>350×/L	137	104		
	<8mg/dL	106	54	13.576	0.001
BUN	8-23mg/dL	5498	4313		
	>23mg/dL	355	340		
	<110mg/dL (	5021	3565	129.170	<0.001
FPG	110-126mg/dL	418	441		
	>126mg/dL	520	701		
	<200mg/dL	4278	3196	22.245	<0.001
TC	200-240mg/dL	1340	1165		
	>240mg/dL	341	346		
	<150mg/dL	4291	2925	121.039	<0.001
TG	≥150and<200mg/dL	796	790		
	≥200mg/dL	872	992		
	<40mg/dL	771	791	31.449	<0.001
HDL-C	≥40mg/dL	5188	3916		
	<130mg/dL	5078	3931	6.275	0.043
LDL-C	≥130and<160mg/dL	699	605		
	≥160mg/dL	182	171		
	0-8mg/L	5695	4383	30.381	<0.001
CRP	>8mg/L	264	324		
	<5%	242	135	75.609	<0.001
HbA1c	5%-8%	5567	4309		
	>8%	150	263		
	<0.59mg/L	398	137	282.423	<0.001
CysC	0.59-1.03mg/L	5081	3739		
	>1.03mg/L	480	831		
	<8.258	1737	937	230.141	<0.001
TyG index	8.258-8.628	1596	1067		
	8.628-9.097	1410	1254		
	>9.097	1216	1449		
	<179.687	1848	818	551.855	<0.001
TyG-BMI	179.687-205.897	1675	992		
index	205.897-235.361	1380	1287		
	>235.361	1056	1610		
	Male<0.6mg/dL, Female				
	<0.5mg/dL	205	126	46.518	<0.001
Cr	Male0.6-1.2mg/dL,				
	Female0.5-1.1mg/dL	5536	4280		

UA	Male>1.2mg/dL, Female >1.1mg/dL	218	301	133.777	<0.001
	Male<3.5mg/dL, Female <2.6mg/dL	280	156		
	Male3.5-7.2mg/dL, Female2.6-6.0mg/dL	5237	3888		
	Male>7.2mg/dL, Female >6.0mg/dL	442	663		
HCT	Male<40%, Female<35%	1056	747	13.030	0.001
	Male40%-50%, Female35%-45%	4450	3529		
	Male>50%, Female>45%	453	431		
	Male<13g/dL, Female< 11.5g/dL	975	625		
Hb	Male13-17.5g/dL, Female11.5-15g/dL	4759	3796	45.915	<0.001
	Male>17.5g/dL, Female> 15g/dL	225	286		

Hyperparameter Optimization

The 17 biochemical indicators identified through univariate analysis were incorporated into four machine learning algorithms to construct cardiovascular disease (CVD) prediction models. To enhance model generalizability, we implemented 5-fold cross-validation on the training set (80% of total data) for hyperparameter tuning, with the area under the receiver operating characteristic curve (AUC) serving as the primary optimization metric. The optimal hyperparameter combinations derived from this process are detailed in Table 2, while Figure 1 visually demonstrates the parameter optimization trajectories across validation folds. This dual approach ensures both rigorous performance benchmarking and transparent reporting of model configuration decisions.

Table 2. Optimized Hyperparameters for Predictive Models

Model	Hyperparameters	Optimal Values
Logistic Regression	Alpha、lambda	Alpha=0.3、lambda=0.00175

RF	mtry、 min_n	Mtry=3、 min_n=5
SVM	Cost、 sigma	Cost=4、 sigma=0.0078125
	max depth、 eta、 gamma	max depth = 3、 eta = 0.1, gamma = 0.2,
XGBOOSToost	colsample bytree、 min child weight、 subsample	colsample bytree = 0.8, min child weight = 3, subsample = 0.6

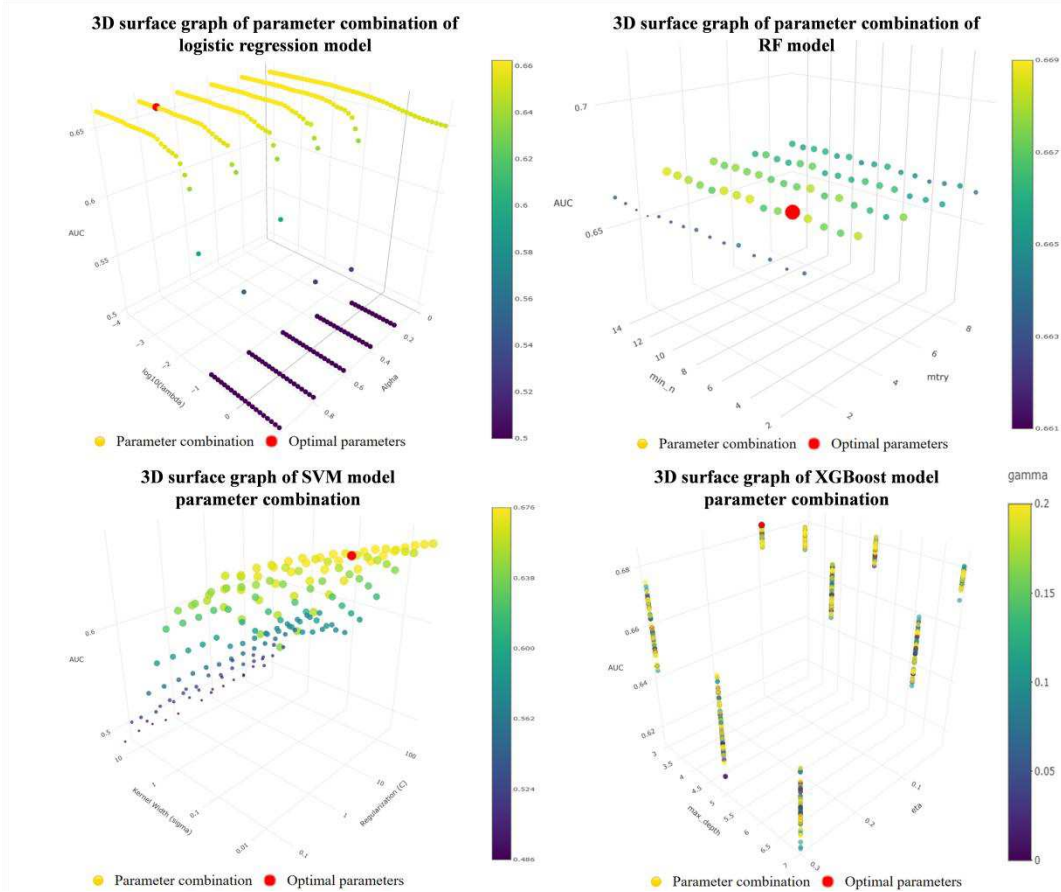


Figure 1. 3D surface plots of parameter combinations for various predictive models

Biomarker Selection

Regularized Logistic Regression

The elastic-net regularized logistic regression model ($\alpha=0.3$, $\lambda=0.00811$) identified significant CVD risk factors through 5,000 bootstrap iterations. Variable importance was quantified via selection frequency (retention rate $\geq 70\%$ indicating high stability), with odds ratios (*OR*) and 95% confidence intervals (*CI*) calculated for significant predictors.

Table 3 Results of the regularized logistic regression model

Index	OR (95% CI)	Bootstrap inclusion rate
WBC Ref= (3.5-9.5×/L)		
(<3.5×/L)	0.895 (0.716 - 1.027)	83.78%
(>9.5×/L)	1.115 (0.974 - 1.391)	83.16%
MCV Ref= (80-100fL)		
(<80fL)	0.870 (0.715 - 1.000)	92.26%
(>100fL)	0.964 (0.844 - 1.067)	71.04%
BUN Ref= (8-23mg/dL)		
(<8mg/dL)	0.825 (0.566 - 1.022)	83.72%
(>23mg/dL)	1.013 (0.891 - 1.171)	63.80%
FPG Ref=Normal (<110mg/dL)		
(<110-126mg/dL)	1.135 (1.000 - 1.326)	94.42%
Diabetes (>126mg/dL)	1.294 (1.111 - 1.514)	99.94%
TC		

Ref= (<200mg/dL)

(200-240mg/dL)	0.993 (0.916 - 1.062)	53.96%
----------------	-----------------------	--------

(>240mg/dL)	1.032 (0.920 - 1.206)	60.34%
-------------	-----------------------	--------

TG

Ref= (<150mg/dL)

(150-200mg/dL)	0.999 (0.910 - 1.100)	57.66%
----------------	-----------------------	--------

(≥200mg/dL)	0.894 (0.778 - 1.000)	94.46%
-------------	-----------------------	--------

HDL-C

Ref= (≥40mg/dL)

(<40mg/dL)	0.973 (0.868 - 1.056)	68.48%
------------	-----------------------	--------

LDL-C

Ref= (<130mg/dL)

(130-160mg/dL)	0.987 (0.880 - 1.081)	60.08%
----------------	-----------------------	--------

(≥160mg/dL)	0.997 (0.827 - 1.194)	55.00%
-------------	-----------------------	--------

CRP

Ref= (0-8mg/L)

(>8mg/L)	1.190 (1.000 - 1.446)	96.38%
----------	-----------------------	--------

HbA1c

Ref= (5%-8%)

(<5%)	0.818 (0.646 - 1.000)	95.88%
(>8%)	1.318 (1.024 - 1.722)	98.26%
CysC Ref= (0.59-1.03mg/L)		
(<0.59mg/L)	0.516 (0.414 - 0.641)	100.00%
(>1.03mg/L)	2.045 (1.775 - 2.359)	100.00%
TyG index Ref=Q1 (<8.258)		
Q2 (8.258-8.628)	0.980 (0.896 - 1.043)	63.86%
Q3 (8.628-9.097)	1.021 (0.963 - 1.119)	61.12%
Q4 (>9.097)	1.003 (0.937 - 1.088)	28.32%
TyG-BMI index Ref=Q1 (<179.687)		
Q2 (179.687-205.897)	1.176 (1.043 - 1.334)	99.52%
Q3 (205.897-235.361)	1.740 (1.541 - 1.970)	100.00%
Q4 (>235.361)	2.732 (2.393 - 3.132)	100.00%
HCT Ref=Normal		

Abnormally low	0.963 (0.863 - 1.014)	69.66%
Abnormally high	0.958 (0.816 - 1.050)	63.52%
Hb Ref=Normal		
Abnormally low	1.043 (0.979 - 1.182)	69.96%
Abnormally high	1.340 (1.069 - 1.696)	99.38%
Cr Ref=Normal		
Abnormally low	0.844 (0.652 - 1.000)	89.70%
Abnormally high	1.186 (1.000 - 1.454)	94.86%
UA Ref=Normal		
Abnormally low	0.984 (0.824 - 1.154)	64.12%
Abnormally high	1.285 (1.106 - 1.490)	99.94%

157 **Random Forest Modeling**

158 The Boruta feature selection algorithm was integrated with Random Forest (RF) modeling to

159 identify cardiovascular risk biomarkers from 17 candidate variables. Through 99 iterations of

160 Boruta optimization, 16 biomarkers demonstrated significant predictive value: TyG index, CysC,

161 HbA1c, TyG-BMI index, TG, FPG, CRP, UA, Cr, Hb, TC, HCT, LDL-C, WBC count, HDL-C,

and MCV(green zone). No variables were classified as tentative (yellow zone), while BUN was rejected as non-predictive (red zone). The feature importance ranking is shown in Figure 2. Selected features were incorporated into an optimized RF model (min_n=5, mtry=3). Dual importance metrics - Mean Decrease Accuracy (clinical relevance) and Mean Decrease Gini (node purity improvement) - identified TyG-BMI and CysC as core predictors. The Boruta-RF hybrid approach enhanced model stability through shadow feature comparison, reducing overfitting risk compared to conventional RF.

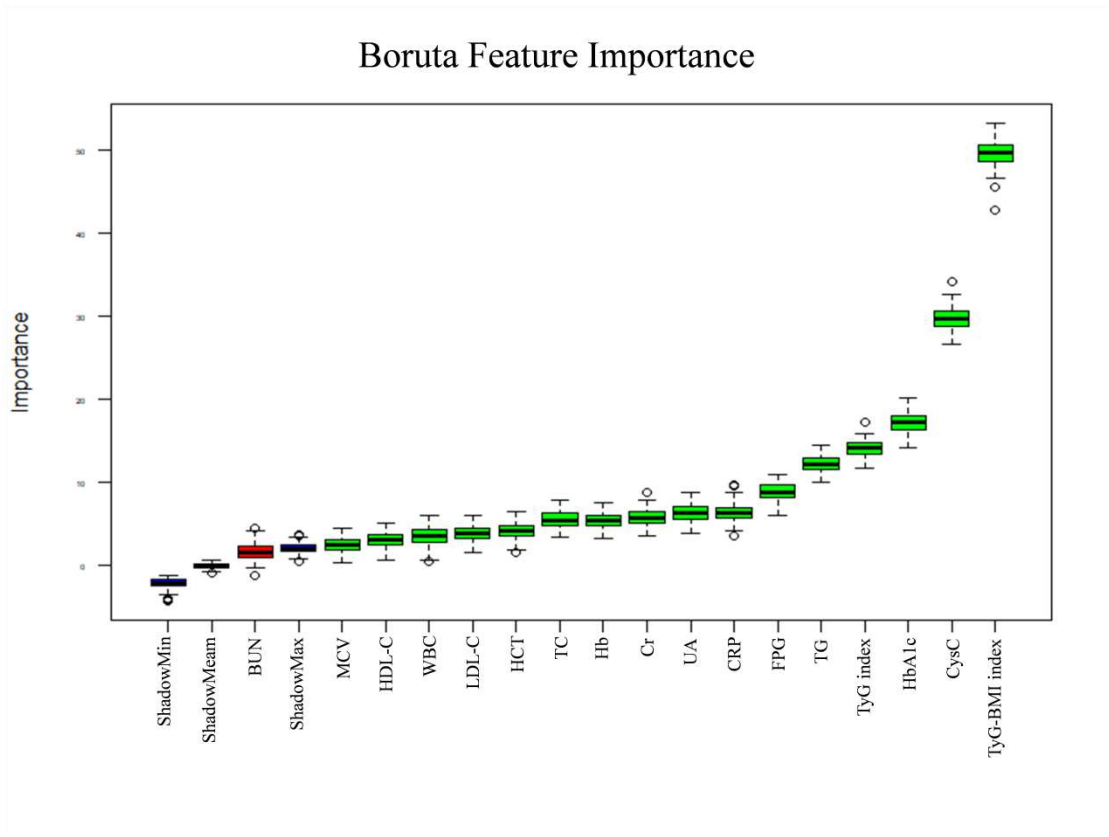


Figure 2. Feature Importance Ranking from Boruta-Enhanced Random Forest Analysis

Random Forest Model Variable Importance

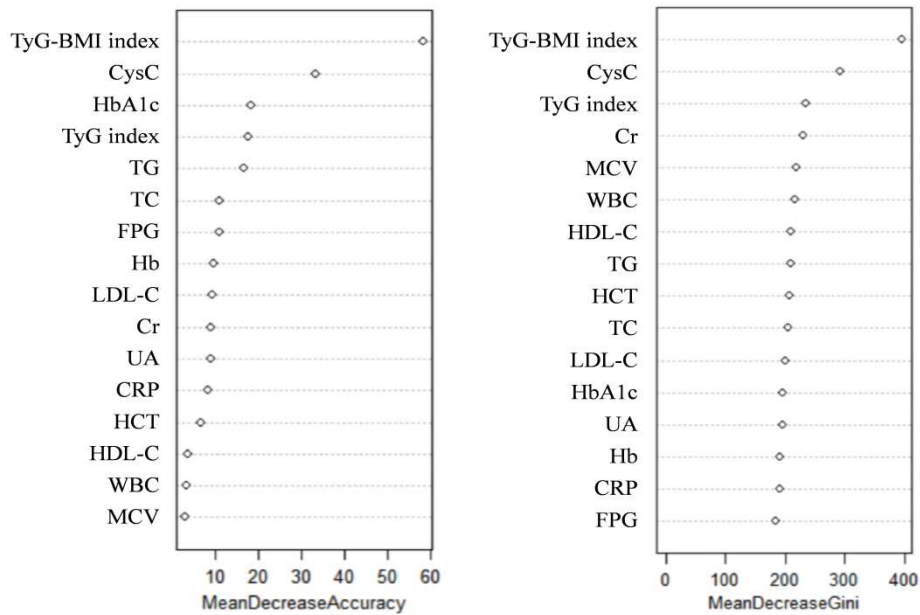


Figure 3. Dual-Perspective Variable Importance in Optimized Random Forest Model

Support Vector Machine Analysis

Using a support vector machine (SVM) ($C=4$, $\gamma=0.0078125$), we developed a cardiovascular disease prediction model through supervised learning. The model achieved a baseline classification accuracy of 65.68% on the independent test set, correctly predicting CVD status for nearly two-thirds of participants. We employed permutation feature importance analysis ($n=500$ iterations) to quantify variable contributions by comparing model accuracy before and after random feature shuffling. Normalized importance scores (range: 0-1) for all 17 biomarkers are ranked in Figure 4, with higher values indicating greater predictive contribution. The analysis identified CysC and the TyG-BMI index as core predictors, while other biomarkers showed substantially lower contributions.

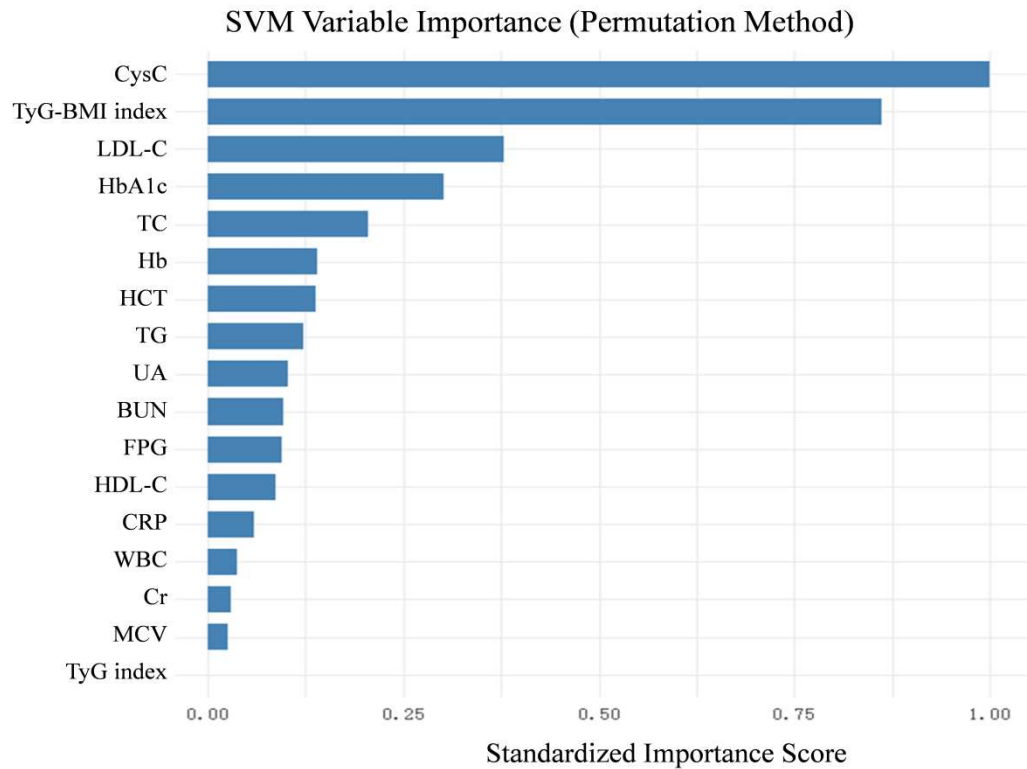


Fig 4 The variable importance of the Support Vector Machine (SVM) Model

XGBoost Modeling

An XGBoost model was optimized with the following hyperparameters: maximum tree depth=3, learning rate (η)=0.1, regularization term (γ)=0.2, column subsampling=0.8, minimum child weight=3, and row subsampling=0.6. Feature importance was quantified using gain score analysis, calculated as: $\text{Gain} = [\text{+} \text{ } - \text{ }] - \gamma$. Higher gain values indicate greater predictive contribution. Hierarchical clustering based on feature co-occurrence patterns revealed two distinct clusters (Figure 5). Cluster 2 contained the TyG-BMI index and CysC, demonstrating their dominant predictive power for CVD risk. Cluster 1 comprised 15 biomarkers with substantially lower gain scores .

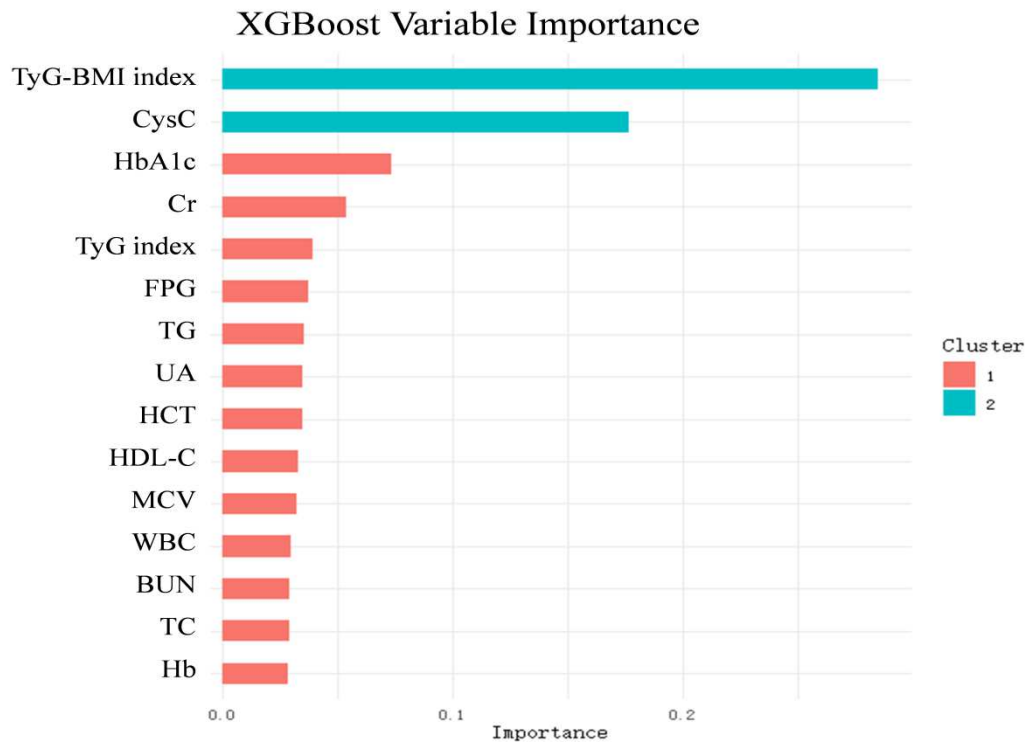


Fig 5 The variable importance of the XGBoost Model

Model Performance Evaluation

Receiver Operating Characteristic Analysis

Model discrimination was assessed using receiver operating characteristic (ROC) curves with area under the curve (AUC) quantification on independent test (20%) and training (80%) sets. On test data, regularized logistic regression achieved superior performance (AUC=0.689), followed by XGBoost (AUC=0.665), random forest (AUC=0.656), and SVM (AUC=0.646). Training set analysis revealed: Random forest showed perfect discrimination (AUC=1.00), indicating severe overfitting, XGBoost maintained robust performance (AUC=0.749), SVM (AUC=0.705) and logistic regression (AUC=0.661) demonstrated appropriate regularization. The random forest model showed the largest performance gap between training and test sets (Difference = -0.344), indicating poor generalizability. In contrast, logistic regression demonstrated the best generalization (Difference = +0.028). Final model rankings: 1. Classification performance: Logistic

regression > XGBoost > Random forest > SVM,2. Generalization ability: Logistic regression > SVM > XGBoost > Random forest.

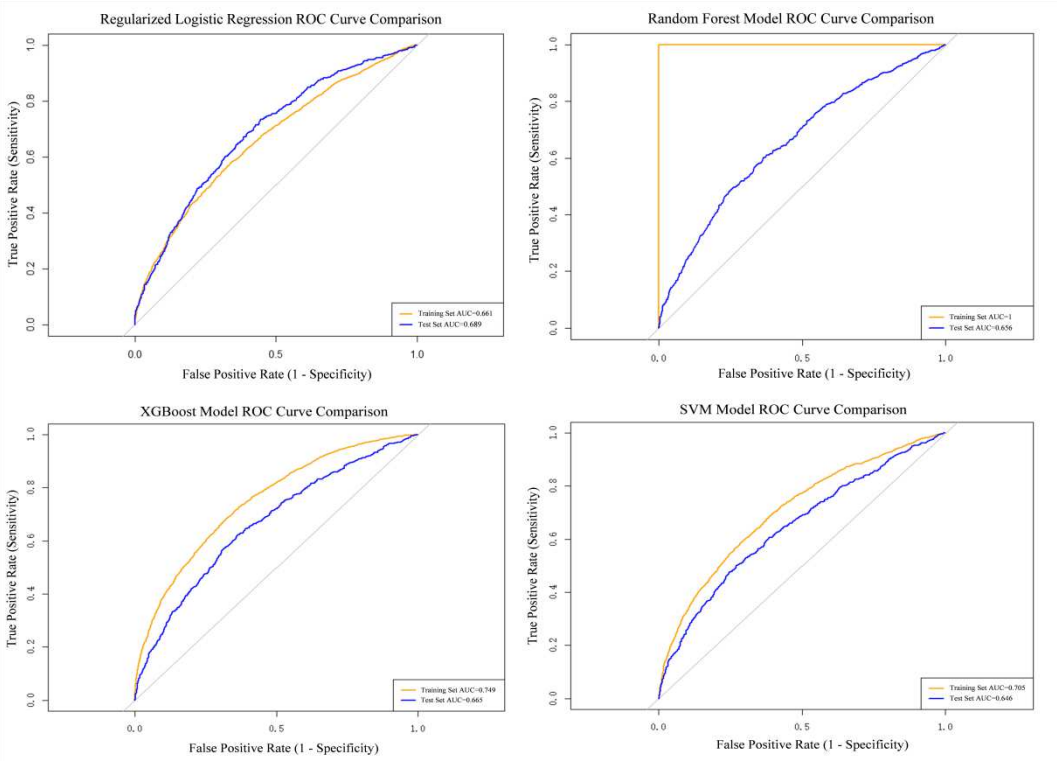


Fig 5 Receiver Operating Characteristic (ROC) curves of the evaluated models

F1-Score Analysis

The logistic regression model showed suboptimal classification performance (test set F1=0.612) with a positive generalization gap (Difference = +0.181), indicating limited generalizability.Random forest demonstrated severe overfitting (training F1=1.00) with the poorest test performance (F1=0.522) and largest generalization gap (Difference = +0.478), confirming it as the least effective model overall.SVM exhibited exceptional generalizability (Difference = +0.063) with test set F1=0.658 outperforming its training performance (F1=0.595).XGBoost achieved superior classification (test F1=0.694) with minimal generalization gap (Difference = -0.045), demonstrating both high accuracy and robustness.

Table 4 The F1 scores of each model on the training and test sets.

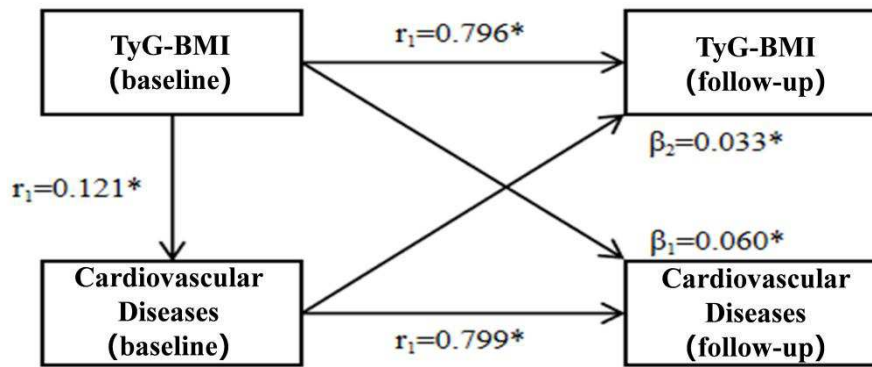
	F1	
	Training set	Test set
Logistic regression	0.372	0.553
RF	1	0.522
SVM	0.537	0.600
XGboost	0.739	0.694

Bidirectional Temporal Analysis Using Cross-Lagged Path Modeling

Through comprehensive evaluation of all four machine learning models, we identified the two most robust cardiovascular disease predictors from the top-performing XGBoost model: the TyG-BMI and CysC. Using matched baseline (2011) and follow-up (2015) cohorts, we conducted cross-lagged path analyses to examine bidirectional temporal relationships between: 1) TyG-BMI and CVD ($n=6,143$), 2) CysC and cardiovascular disease CVD ($n=5,213$).

Bidirectional Temporal Relationships Between TyG-BMI and Cardiovascular Disease

Given significant variance in TyG-BMI measurements, we performed Z-score normalization prior to analysis (mean=0, SD=1). All models adjusted for baseline (2011) and follow-up (2015) covariates including: 1) Sex, 2) Residential location, 3) Education level, 4) Age categories. TyG-BMI demonstrated strong temporal stability (autoregressive coefficient=0.796, 95% CI: 0.781-0.811, $P<0.001$), Cardiovascular disease CVD showed similar stability (autoregressive coefficient=0.799, 95% CI: 0.782-0.816, $P<0.001$). Baseline TyG-BMI correlated significantly with baseline CVD status ($r=0.121$, 95% CI: 0.109-0.133, $P<0.001$), the cross-lagged path from baseline TyG-BMI to follow-up CVD showed significant association ($\beta=0.060$, $P<0.001$), the reverse pathway (baseline CVD $\rightarrow$ follow-up TyG-BMI) was also significant ($\beta=0.033$, $P<0.001$). Path analysis confirmed bidirectional causality between TyG-BMI and cardiovascular disease CVD (Figure 7).



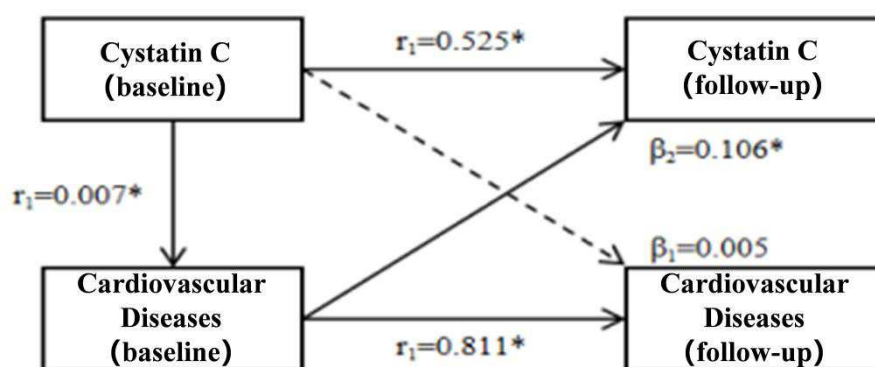
* indicates statistically significant coefficients ($P < 0.05$), solid lines represent statistically significant associations ($P < 0.05$), dashed lines indicate non-significant associations ($P \geq 0.05$).

Fig 7 TyG-BMI Index and Cardiovascular Diseases: Cross-Lagged Path Diagram

* indicates statistically significant coefficients ($P < 0.05$), solid lines represent statistically significant associations ($P < 0.05$), dashed lines indicate non-significant associations ($P \geq 0.05$).

Temporal Relationships Between Cystatin C and Cardiovascular Disease

After adjusting for baseline (2011) and follow-up (2015) covariates including sex, residence, education level, and age categories, cCystatin C showed moderate temporal stability (autoregressive coefficient=0.525, 95% CI: 0.503-0.547, $P < 0.001$), while cardiovascular diseaseCVD demonstrated stronger stability (autoregressive coefficient=0.811, 95% CI: 0.793-0.829, $P < 0.001$). The baseline correlation between Ccystatin C and CVD was significant but minimal ($r = 0.007$, 95% CI: 0.004-0.010, $P < 0.001$). The cross-lagged path from baseline CysC to follow-up CVD was non-significant ($\beta = 0.005$, $P = 0.585$), whereas the reverse pathway showed significant association (baseline CVD $\rightarrow$ follow-up CysC: $\beta = 0.106$, $P < 0.001$). Path analysis revealed unidirectional causality from CVD to CysC (Figure 8).



* indicates statistically significant coefficients ($P < 0.05$), solid lines represent statistically significant associations ($P < 0.05$), dashed lines indicate non-significant associations ($P \geq 0.05$).

Fig 8 Cystatin C and Cardiovascular Diseases: Cross-Lagged Path Diagram

Discussion

This study pioneers a novel biomarker screening framework by integrating machine learning algorithms with cross-lagged path modeling for cardiovascular CVD research. Compared to conventional single-model approaches, our multi-algorithm comparison enhanced robustness in variable selection while mitigating overfitting and selection bias inherent to individual models. The cross-lagged path analysis provided longitudinal evidence for temporal relationships between metabolic biomarkers and CVD progression over 4-year follow-up. This integrated multi-algorithm, multi-model paradigm establishes a new methodological standard for investigating multidimensional etiology in complex chronic diseases. We identified a substantially higher CVD prevalence among Chinese adults aged ≥ 45 years (44.13%, 4707/10666) compared to European¹³ and American¹⁴ counterparts. This disparity likely reflects synergistic effects of: 1) High metabolic syndrome prevalence, 2) Accelerated population aging, 3) Urban-rural healthcare resource disparities. China's metabolic syndrome prevalence ranks among the world's highest at 31.1% in adults ≥ 20 years (affecting >450 million)¹⁵. The rising trend¹⁶ and synergistic metabolic

effects accelerate atherosclerotic progression¹⁷, explaining elevated CVD risks in China's aging population. Rural CVD mortality remains 1.19-fold higher than urban areas (2023 data)¹⁸ underscoring primary care deficiencies in early screening and intervention. Rapid demographic aging (12 million additional seniors annually, 2010-2020)¹⁹ compounds CVD risks, as elderly face higher incidence than younger adults²⁰. Among four tested models (regularized logistic regression, RF, SVM, XGBoost), XGBoost demonstrated superior biomarker selection performance, identifying TyG-BMI and CysC as core predictive indicators.

The TyG-BMI is a composite metabolic indicator that integrates measures of insulin resistance (TyG index), obesity (BMI), and glycemic control. Previous studies^{21 22 23} have demonstrated TyG-BMI's significant predictive value for both CVD incidence and mortality, which our multi-model analysis confirms as a core predictor. Regularized logistic regression showed each index increase in TyG-BMI was associated with higher CVD risk. These findings align with previous reports by Chuka et al.²⁴ and RUN et al.²⁵, and are further supported by a Chinese community-based study revealed that elevated TyG-BMI levels significantly predicted increased 10-year risks of cardiovascular events²⁶. Cross-lagged path analysis revealed high temporal stability for both TyG-BMI and CVD, this suggests that once elevated TyG-BMI and CVD status are established, they become relatively resistant to change. These findings underscore the critical importance of early screening and preventive interventions for CVD risk reduction²⁷. Furthermore, we identified bidirectional causality between TyG-BMI and CVD. This likely reflects a vicious cycle between metabolic dysregulation and cardiovascular risk progression. Our study focused on middle-aged and elderly populations, where aging-related and CVD-induced cardiovascular homeostasis deterioration shows strong associations with metabolic disorders (e.g., insulin

296 resistance)²⁸.CVD patients typically exhibit metabolic alterations characterized by chronic
297 low-grade inflammation and metaflammation ²⁹ ,These inflammatory states can trigger
298 hypercholesterolemia and obesity³⁰,subsequently exacerbating metabolic dysregulation in CVD
299 patients.Metabolic syndrome - a cluster of metabolic abnormalities including insulin resistance,
300 abdominal obesity, and dyslipidemia - is mechanistically linked to elevated oxidative stress and
301 inflammation ³¹ ,these pathological processes represent key drivers of CVD
302 pathogenesis³².Furthermore, metabolic disorders may induce CVD through impaired cardiac
303 systolic/diastolic function³³.Mitochondrial dysfunction serves as another critical nexus between
304 metabolic disorders and CVD,where impaired oxidative phosphorylation and mitochondrial
305 biogenesis contribute to both metabolic syndrome progression and cardiovascular
306 pathogenesis³⁴,suggesting shared subcellular pathological foundations.Cystatin C, a sensitive
307 biomarker of glomerular filtration rate,demonstrates robust associations with coronary artery
308 disease, stroke, and cardiovascular events³⁵,and shows excellent predictive value for CVD
309 incidence³⁶³⁷.Our multi-model analysis identified CysC as a core CVD predictor across all
310 algorithms.Regularized logistic regression demonstrated that elevated CysC more than doubled
311 CVD risk.These findings align with Yanjun et al.'s report of CysC's predictive value in Chinese
312 populations³⁸,and are further supported by evidence linking elevated CysC to cardiovascular
313 mortality ³⁹ ,notably, this association persists even in populations with normal renal
314 function⁴⁰.Autoregressive analysis revealed moderate temporal stability for CysC ($\beta=0.525$)
315 versus high stability for CVD ($\beta=0.811$),the significantly lower stability coefficient for CysC
316 compared to TyG-BMI (0.525 vs. 0.796),suggests greater susceptibility of renal markers to disease
317 progression and therapeutic modulation than metabolic indices.The

cardiovascular-kidney-metabolic (CKM) syndrome frequently co-occur as interrelated conditions. Patients with CKM exhibit significantly higher risks of both acute kidney injury and chronic kidney disease, sharing common pathophysiological pathways⁴¹. Cross-lagged analysis established unidirectional causality from CVD to CysC elevation, suggesting cardiovascular pathology precedes renal biomarker abnormalities. Heart failure - a major renal risk factor - induces acute/chronic kidney dysfunction through fluid overload and venous congestion mechanisms⁴². Conversely, renal impairment independently elevates CVD risk⁴³, creating a vicious cycle of cardio-renal dysfunction that perturbs biomarker levels. Furthermore, CysC levels demonstrate significant positive correlations with inflammatory markers⁴⁴, and since systemic inflammation serves as a pivotal driver of CVD pathogenesis⁴⁵, this inflammatory cascade may secondarily modulate CysC concentrations through cardio-renal axis activation.

Conclusion

Based on our findings, we propose the following clinical recommendations: 1) Implement routine cardiovascular screening for high-risk populations (middle-aged/elderly adults, obese individuals) with personalized interventions targeting their specific metabolic profiles. 2) Prioritize renal function monitoring in CVD patients to prevent secondary renal impairment or cardio-renal comorbidity. Study limitations include the predominantly middle-aged/elderly cohort, necessitating future validation of model generalizability in younger populations. Furthermore, the causal relationships require verification through Mendelian randomization studies or interventional trials.

Abbreviations

CHARLS China Health and Retirement Longitudinal Study

CVD Cardiovascular diseases

340	BMI	Body mass index
341	HDL-C	Low-density lipoprotein cholesterol
342	LDL-C	Low-density lipoprotein cholesterol
343	OR	Odds ratio
344	TC	Total cholesterol
345	TG	Triglyceride
346	HCT	Hematocrit
347	PLT	Platelet
348	CysC	Cystatin C
349	UA	Uric acid
350	Cr	Creatinine
351	Hb	Hemoglobin
352	FPG	Fasting plasma glucose
353	IFG	Impaired fasting glucose
354	WBC	white blood cell
355	BUN	Blood Urea Nitrogen
356	CRP	C-reactive protein
357	HbA1c	Hemoglobin A1c
358	MCV	Mean Corpuscular Volume
359	TyG	Triglyceride-glucose
360	TyG-BMI	Triglyceride-glucose-BMI
361	ROC	Receiver operating characteristic

362 AUC Area under the curve

363 RF Random forest

364 SVM Support vector machine

365 **Acknowledgements**

366 We sincerely appreciate the efforts and contributions of all participants in the CHALRS project.

367 **Authors contributions**

368 Mingdong Zhang: Conceptualized and designed the study, extracted and organized data,

369 participated in partial data analysis, and contributed to manuscript writing.Yapeng Huang:

370 Organized data and participated in partial data analysis.Zhaoding Liang: Participated in partial

371 data analysis, wrote sections of the manuscript, and created some figures.Pei Liu: Contributed to

372 manuscript writing, created some figures, and reviewed/edited the manuscript. Gaofeng Sun:

373 Writing – review & editing, Conceptualization, Data curation, Funding acquisition.These

374 authors have contributed equally to this work and share last authorship. All authors reviewed and

375 approved the manuscript.

376 **Funding**

377 This research was supported by the National Key Research and Development Program of China

378 (2017YFC0907203), The National Natural Science Foundation of China (U1803124) and

379 Xinjiang Uyghur Autonomous"Tianshan Talent" High-Level Talent Training Program in Medicine

380 and Health (TSYC202301B073).

381 **Data Availability**

382 The data used for these analyses are all publicly available at online(<http://charls.pku.edu.cn/en>).

383 **Declarations**

Ethics approval and consent to participate

The China Health and Retirement Longitudinal Study (CHARLS) database utilized in this research was approved and permitted by the Peking University Institutional Review Board (IRB) (Approval No. IRB00001052-11015). Written informed consent was obtained from all participants involved in the surveys.

Consent for publication

All authors gave their consent for publication.

Competing interests

The authors declare that they have no competing interest.

Reference

- 1 Li Z, Lin L, Wu H, *et al.* Global, Regional, and National Death, and Disability-Adjusted Life-Years (DALYs) for Cardiovascular Disease in 2017 and Trends and Risk Analysis From 1990 to 2017 Using the Global Burden of Disease Study and Implications for Prevention. *Front Public Health*. 2021;9:559751. Published 2021 Oct 29. doi:10.3389/fpubh.2021.559751
- 2 Wang L, Wang Y, Ji Z, *et al.* Trends of cause-specific cardiovascular disease mortality in China, 2009-2019: a nationwide longitudinal study. *Ann Med*. 2025;57(1):2455534. doi:10.1080/07853890.2025.2455534
- 3 Wang Z, Ma L, Liu M, *et al.* The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Summary of the 2022 Report on Cardiovascular Health and Diseases in China. *Chin Med J (Engl)*. 2023;136(24):2899-2908. (In Chinese). doi:10.1097/CM9.0000000000002927
- 4 Goh RSJ, Chong B, Jayabaskaran J, *et al.* The burden of cardiovascular disease in Asia from 2025 to 2050: a forecast analysis for East Asia, South Asia, South-East Asia, Central Asia, and high-income Asia Pacific regions. *Lancet Reg Health West Pac*. 2024;49:101138. Published 2024 Jul 10. doi:10.1016/j.lanwpc.2024.101138
- 5 Lv B, Song G, Jing F, *et al.* Mortality from cerebrovascular diseases in China: Exploration of recent and future trends. *Chin Med J (Engl)*. 2024;137(5):588-595. doi:10.1097/CM9.0000000000002760

-
- 6 Wang J, Wang Z, Guo F, *et al.* Individual and Combined Cardiometabolic Morbidities and the Subsequent Risk of Cardiovascular Events in Chinese Adults. *J Clin Endocrinol Metab.* 2022;107(1):e84-e94. doi:10.1210/clinem/dgab609
 - 7 Hedayatnia M, Asadi Z, Zare-Feyzabadi R, *et al.* Dyslipidemia and cardiovascular disease risk among the MASHAD study population. *Lipids Health Dis.* 2020;19(1):42. Published 2020 Mar 16. doi:10.1186/s12944-020-01204-y
 - 8 Visseren FLJ, Mach F, Smulders YM, *et al.* 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice [published correction appears in Eur Heart J. 2022 Nov 7;43(42):4468. doi: 10.1093/eurheartj/ehac458.]. *Eur Heart J.* 2021;42(34):3227-3337. doi:10.1093/eurheartj/ehab484
 - 9 Sabatine MS, Giugliano RP, Keech A, *et al.* Rationale and design of the Further cardiovascular Outcomes Research with PCSK9 Inhibition in subjects with Elevated Risk trial. *Am Heart J.* 2016;173:94-101. doi:10.1016/j.ahj.2015.11.015
 - 10 Zhu J, Gao R, Zhao S, *et al.* Guidelines for the Prevention and Treatment of Dyslipidemia in Chinese Adults (2016 Revision) . *Chinese Circulation Journal*, 2016, 31(10): 937-953. (In Chinese). doi: 10.3969/j.issn.1000-3614.2016.10.001
 - 11 Shang H, Wang Y, Shen Z. National Guide to Clinical Laboratory Procedures (4th Edition) . People's Medical Publishing House, Beijing, 2015, 3: P196, P629-790.(In Chinese)
 - 12 Chinese Diabetes Society. Chinese Guidelines for the Prevention and Treatment of Diabetes (2024 Edition). *Chinese Journal of Diabetes*, 2025, 17(1): 16-139. (In Chinese). doi:10.3760/cma.j.cn115791-20241203-00705.
 - 13 Timmis A, Vardas P, Townsend N, *et al.* European Society of Cardiology: cardiovascular disease statistics 2021 [published correction appears in Eur Heart J. 2022 Feb 22;43(8):799. doi: 10.1093/eurheartj/ehac064.]. *Eur Heart J.* 2022;43(8):716-799. doi:10.1093/eurheartj/ehab892
 - 14 Zhang J, Fang Y, Yao Y, *et al.* Disparities in cardiovascular disease prevalence among middle-aged and older adults: Roles of socioeconomic position, social connection, and behavioral and physiological risk factors. *Front Cardiovasc Med.* 2022;9:972683. Published 2022 Oct 14. doi:10.3389/fcvm.2022.972683
 - 15 Endocrinology and Metabolic Diseases Committee of Chinese Association of Integrative Medicine Practitioners, Ni Q. Guidelines for Syndrome-Disease Combined Diagnosis and Treatment of Metabolic Syndrome. *World Chinese Medicine*, 2023, 18(22): 3157-3166. (In Chinese). doi:10.3969/j.issn.1673-7202.2023.22.001
 - 16 Ge H, Yang Z, Li X, *et al.* The prevalence and associated factors of metabolic syndrome in Chinese aging population. *Sci Rep.* 2020;10(1):20034. Published 2020 Nov 18. doi:10.1038/s41598-020-77184-x
 - 17 Poznyak A, Grechko AV, Poggio P, *et al.* The Diabetes Mellitus-Atherosclerosis Connection: The Role of Lipid and Glucose Metabolism and Chronic Inflammation. *Int J Mol Sci.* 2020;21(5):1835. Published 2020 Mar 6.

doi:10.3390/ijms21051835

- 18 Liu M, He X, Yang X, *et al.* Key Interpretation of the Report on Cardiovascular Health and Diseases in China 2023 . *Chinese Journal of Cardiovascular Medicine*, 2024, 29(04): 305-324. (In Chinese). doi:10. 3969 / j. issn. 1007-5410. 2024. 04. 002
- 19 Huang, Y. Special issue: challenges of population ageing in China. *China Economic Journal*. 2020;13(1):1-2. doi: 10.1080/17538963.2019.1710058
- 20 Rodgers JL, Jones J, Bolleddu SI, *et al.* Cardiovascular Risks Associated with Gender and Aging. *J Cardiovasc Dev Dis*. 2019;6(2):19. Published 2019 Apr 27. doi:10.3390/jcdd6020019
- 21 Zheng D, Cai J, Xu S, *et al.* The association of triglyceride-glucose index and combined obesity indicators with chest pain and risk of cardiovascular disease in American population with pre-diabetes or diabetes. *Front Endocrinol (Lausanne)*. 2024;15:1471535. Published 2024 Sep 6. doi:10.3389/fendo.2024.1471535
- 22 Park HM, Han T, Heo SJ, *et al.* Effectiveness of the triglyceride-glucose index and triglyceride-glucose-related indices in predicting cardiovascular disease in middle-aged and older adults: A prospective cohort study. *J Clin Lipidol*. 2024;18(1):e70-e79. doi:10.1016/j.jacl.2023.11.006
- 23 Sun Y, Hu Y. Association of triglyceride-glucose-body mass index with all-cause mortality among individuals with cardiovascular disease: results from NHANES. *Front Endocrinol (Lausanne)*. 2025;16:1529004. Published 2025 Jan 27. doi:10.3389/fendo.2025.1529004
- 24 Ononye C, Bapat S, Chebrolu S, *et al.* Abstract 4147006: Relationship between Triglyceride-Glucose Index, Subclinical Myocardial Injury, and Cardiovascular Mortality in NHANES III. *Circulation*, 2024, 150(Suppl_1): A4147006-A4147006. doi:10.1161/circ.150.suppl_1.4147006
- 25 Wang R, Cheng X, Tao W. Association between triglyceride glucose body mass index and cardiovascular disease in adults: evidence from NHANES 2011- 2020. *Front Endocrinol (Lausanne)*. 2024;15:1362667. Published 2024 Jul 16. doi:10.3389/fendo.2024.1362667
- 26 Luo X Y , Fan F F , Jia J , *et al.* The associations of triglyceride glucose and its derived indexes with 10-year cardiovascular disease risk in a Chinese community-based population. *European Heart Journal*, 2024(Supplement_1):Supplement_1. doi:10.1093/eurheartj/ehae666.2752.
- 27 Flores Soler J, Vila Costa I, Bengoa Terrero C, *et al.* Prevecardio: A single-visit population screening model for detection and prevention of 10-year cardiovascular risk. *European Heart Journal*, 2024, 45(Supplement_1). *European Heart*

- 28 Costantino S, Paneni F, Cosentino F. Ageing, metabolism and cardiovascular disease. *J Physiol.* 2016;594(8):2061-2073. doi:10.1113/JP270538
- 29 Silveira Rossi JL, Barbalho SM, Reverete de Araujo R, *et al*, Sloan LA. Metabolic syndrome and cardiovascular diseases: Going beyond traditional risk factors. *Diabetes Metab Res Rev.* 2022;38(3):e3502. doi:10.1002/dmrr.3502
- 30 Domingo E, Marques P, Francisco V, *et al*. Targeting systemic inflammation in metabolic disorders. A therapeutic candidate for the prevention of cardiovascular diseases?. *Pharmacol Res.* 2024;200:107058. doi:10.1016/j.phrs.2024.107058
- 31 Monserrat-Mesquida M, Quetglas-Llabrés M, Capó X, *et al*. Metabolic Syndrome is Associated with Oxidative Stress and Proinflammatory State. *Antioxidants (Basel).* 2020;9(3):236. Published 2020 Mar 12. doi:10.3390/antiox9030236
- 32 Steven S, Frenis K, Oelze M, *et al*. Vascular Inflammation and Oxidative Stress: Major Triggers for Cardiovascular Disease. *Oxid Med Cell Longev.* 2019;2019:7092151. Published 2019 Jun 23. doi:10.1155/2019/7092151
- 33 Iacoviello M. Metabolic Regulation in the Development of Cardiovascular Disease and Heart Failure. *Int J Mol Sci.* 2023;24(10):8773. Published 2023 May 15. doi:10.3390/ijms24108773
- 34 Ren J, Pulakat L, Whaley-Connell A, *et al*. Mitochondrial biogenesis in the metabolic syndrome and cardiovascular disease. *J Mol Med (Berl).* 2010;88(10):993-1001. doi:10.1007/s00109-010-0663-9
- 35 van der Laan SW, Fall T, Soumaré A, *et al*. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *J Am Coll Cardiol.* 2016;68(9):934-945. doi:10.1016/j.jacc.2016.05.092
- 36 Plattner T, Larcher B, Mader A, *et al*. Cystatin c predicts cardiovascular events in patients with coronary artery disease both among patients with and among those without type 2 diabetes[J]. *European Heart Journal*, 2024, 45(Supplement_1).
- 37 Mader A, Maechler M, Larcher B, *et al*. 157-or: Cystatin c predicts major cardiovascular events in patients with coronary artery disease both among patients with and among those without the metabolic syndrome[J]. *Diabetes*, 2023, 72(Supplement_1). <https://doi.org/10.2337/db23-157-OR>
- 38 Zhang Y, Yang S, Chen J, *et al*. Associations of serum cystatin C and its change with new-onset cardiovascular disease in Chinese general population. *Nutr Metab Cardiovasc Dis.* 2022;32(8):1963-1971. doi:10.1016/j.numecd.2022.05.016
- 39 Song X, Xiong L, Guo T, *et al*. Cystatin C Is a Predictor for Long-Term, All-Cause, and Cardiovascular Mortality in US Adults With Metabolic Syndrome. *J Clin Endocrinol Metab.* 2024;109(11):2905-2919. doi:10.1210/clinem/dgae225
- 40 Veeranna V, Zalawadiya S K, Bavishi C, *et al*. Abstract 17860: Association between cystatin c and red cell distribution width with cardiovascular mortality among a nationally representative population and normal renal function. *Circulation*,

2012, 126(suppl_21): A17860-A17860. doi:10.1161/circ.126.suppl_21.A17860

- 41 Kanbay M, Guldan M, Ozbek L, *et al.* Exploring the nexus: The place of kidney diseases within the cardiovascular-kidney-metabolic syndrome spectrum. *Eur J Intern Med.* 2024;127:1-14. doi:10.1016/j.ejim.2024.07.014
- 42 Deferrari G, Cipriani A, La Porta E. Renal dysfunction in cardiovascular diseases and its consequences. *J Nephrol.* 2021;34(1):137-153. doi:10.1007/s40620-020-00842-w
- 43 Lv Z, Fu Y, Liu C, *et al.* The role of cardiac remodeling associated with renal function in mediating cardiovascular event outcomes. *iScience.* 2024;27(3):109143. Published 2024 Feb 5. doi:10.1016/j.isci.2024.109143
- 44 Zou J, Sun H, Xiang Y. Correlation of serum cystatin C with inflammatory cytokines in patients with traumatic brain injury. *Synapse.* 2021;75(8):e22201. doi:10.1002/syn.22201
- 45 Henein MY, Vancheri S, Longo G, *et al.* The Role of Inflammation in Cardiovascular Disease. *Int J Mol Sci.* 2022;23(21):12906. Published 2022 Oct 26. doi:10.3390/ijms232112906