

Supplementary Materials

Predicting cardiovascular-kidney-metabolic multimorbidity in Chinese adults with overweight and obesity using machine learning: an internal evaluation

Xin Li^{1,#}, Zhitong Li^{1,#}, Yexin Ji¹, Yuanyuan Yan¹, Xiaoqin Chen^{1,2}, Linlin Gao^{1,2}, Baoling Wei^{1,2}, Chenxia Zhang^{1,2}, Ruixue Duan^{1,2}, Shiwei Liu^{1,2}

¹Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Taiyuan 030032, Shanxi, China.

²Department of Endocrinology, Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Taiyuan 030032, Shanxi, China.

[#]These authors contributed equally to this work and share first authorship.

Correspondence to: Prof. Shiwei Liu, Department of Endocrinology, Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Taiyuan 030032, Shanxi, China. E-mail: liushiwei@sxbqeh.com.cn

Supplementary Table 1. Definition of cardiovascular-kidney-metabolic multimorbidity in the China Health and Retirement Longitudinal Study (CHARLS)

Variables	Questionnaire	Definition
Cardiovascular conditions	● whether the respondent reported having heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems. ● whether the respondent reported having a stroke.	1: Having any of them 0: Having none of them
Kidney condition	whether the respondent reported having any kidney disease (except for tumor or cancer).	1: Yes 0: No
Metabolic condition	● whether the respondent reported having hypertension. ● whether the respondent reported having diabetes or high blood sugar. ● whether the respondent reported ever having dyslipidemia (elevation of low density lipoprotein, triglycerides (TGs), and total cholesterol, or a low high density lipoprotein level).	1: Having any of them 0: Having none of them
Variables	Subtype	Definition
Cardiovascular-kidney-metabolic (CKM) multimorbidity	Cardiovascular-kidney (CK) multimorbidity	C=1 & K=1 &
	Cardiovascular-metabolic (CM) multimorbidity	M=0
	Kidney-metabolic (KM) multimorbidity	C=1 & K=0 & M=1
	Complete CKM multimorbidity	C=0 & K=1 & M=1
		C=1 & K=1 & M=1

Supplementary Table 2. Optimal hyperparameters for six machine learning models

Algorithms	Optimal hyperparameters
Decision Tree (DT)	ccp_alpha: 0.0
	max_depth: 5
	max_features: sqrt
	min_samples_leaf: 1
	min_samples_split: 20
Random Forest (RF)	n_estimators = 200
	max_features = 3
	learning_rate: 0.01
XGBoost	max_depth: 3
	n_estimators: 200
	subsample: 0.6
	colsample_bytree: 0.6
LightGBM	learning_rate: 0.1
	n_estimators: 100
	num_leaves: 31
	subsample: 0.6
Support Vector Machine (SVM)	C: 0.1
	degree: 2
	gamma: scale
	kernel: linear
Artificial Neural Networks (ANN)	activation: logistic
	hidden_layer_sizes: (100,)

Supplementary Table 3. Baseline characteristics of the participants

	Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>value</i>
Age (median [IQR]), years	58.00 [52.00, 64.00]	57.00 [51.00, 64.00]	61.00 [55.00, 67.00]	<0.0 01
Sex (%)				0.35 8
Male	2,521 (59.40)	2,125 (59.09)	396 (61.11)	
Female	1,723 (40.60)	1,471 (40.91)	252 (38.89)	
Region (%)				<0.0 01
Central	1,553 (36.59)	1,274 (35.43)	279 (43.06)	
East	1,622 (38.22)	1,413 (39.29)	209 (32.25)	
West	1,069 (25.19)	909 (25.28)	160 (24.69)	
Education (%)				0.01 5
Less than elementary school	1,757 (41.40)	1,459 (40.57)	298 (45.99)	
Elementary school	925 (21.80)	795 (22.11)	130 (20.06)	
Middle school	1,017 (23.96)	887 (24.67)	130 (20.06)	
High school or above	545 (12.84)	455 (12.65)	90 (13.89)	
Marital (%)				0.18 8
Married	3,921 (92.39)	3,331 (92.63)	590 (91.05)	
Others	323 (7.61)	265 (7.37)	58 (8.95)	
Residence (%)				0.03 6
Rural	2,622 (61.78)	2,246 (62.46)	376 (58.02)	
Urban	1,622 (38.22)	1,350 (37.54)	272 (41.98)	
Living Standard (%)				0.48 2
Poor	505 (11.90)	431 (11.99)	74 (11.42)	

	Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>valu</i> <i>e</i>
Relatively poor	847 (19.96)	707 (19.66)	140 (21.60)	
Average	2,582 (60.84)	2,198 (61.12)	384 (59.26)	
Relatively high	261 (6.15)	222 (6.17)	39 (6.02)	
Very high	49 (1.15)	38 (1.06)	11 (1.70)	
Socioeconomic status (%)				0.37 6
Low	827 (19.49)	687 (19.10)	140 (21.60)	
Low-middle	2,150 (50.66)	1,822 (50.67)	328 (50.62)	
Upper-middle	1,206 (28.42)	1,036 (28.81)	170 (26.23)	
High	61 (1.44)	51 (1.42)	10 (1.54)	
Social Isolation (%)				0.92 2
No	3,062 (72.15)	2,596 (72.19)	466 (71.91)	
Yes	1,182 (27.85)	1,000 (27.81)	182 (28.09)	
Smoking (%)				0.32
No	2,703 (63.69)	2,302 (64.02)	401 (61.88)	
Yes	1,541 (36.31)	1,294 (35.98)	247 (38.12)	
Drinking (%)				0.38 4
No	2,875 (67.74)	2,426 (67.46)	449 (69.29)	
Yes	1,369 (32.26)	1,170 (32.54)	199 (30.71)	
Sleep Duration (%)				0.00 6
≤6 h	2,044 (48.16)	1,702 (47.33)	342 (52.78)	
6–8 h	1,789 (42.15)	1,553 (43.19)	236 (36.42)	
>8 h	411 (9.68)	341 (9.48)	70 (10.80)	
MET (median [IQR]), min	4,158.00 [462.00, 9,198.00]	4,158.00 [462.00, 9,600.00]	2,880.00 [80.00, 7,231.00]	<0.0 01

		Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>value</i>
Depression (%)					<0.001
	No	2,222 (52.36)	1,967 (54.70)	255 (39.35)	
	Yes	2,022 (47.64)	1,629 (45.30)	393 (60.65)	
Dementia (%)					0.059
	No	4,035 (95.08)	3,429 (95.36)	606 (93.52)	
	Yes	209 (4.92)	167 (4.64)	42 (6.48)	
Dyslipidemia (%)					<0.001
	No	3,497 (82.40)	3,062 (85.15)	435 (67.13)	
	Yes	747 (17.60)	534 (14.85)	213 (32.87)	
Hypertension (%)					<0.001
	No	2,196 (51.74)	1,991 (55.37)	205 (31.64)	
	Yes	2,048 (48.26)	1,605 (44.63)	443 (68.36)	
Diabetes (%)					<0.001
	No	3,616 (85.20)	3,112 (86.54)	504 (77.78)	
	Yes	628 (14.80)	484 (13.46)	144 (22.22)	
Heart Disease (%)					<0.001
	No	4,008 (94.44)	3,440 (95.66)	568 (87.65)	
	Yes	236 (5.56)	156 (4.34)	80 (12.35)	
Stroke (%)					0.004
	No	4,217 (99.36)	3,579 (99.53)	638 (98.46)	
	Yes	27 (0.64)	17 (0.47)	10 (1.54)	
Liver Disease (%)					<0.001

		Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>valu</i> <i>e</i>
	No	4,055 (95.55)	3,459 (96.19)	596 (91.98)	
	Yes	189 (4.45)	137 (3.81)	52 (8.02)	
Kidney Disease (%)					<0.0 01
	No	4,124 (97.17)	3,514 (97.72)	610 (94.14)	
	Yes	120 (2.83)	82 (2.28)	38 (5.86)	
Digestive Disease (%)					<0.0 01
	No	3,140 (73.99)	2,707 (75.28)	433 (66.82)	
	Yes	1,104 (26.01)	889 (24.72)	215 (33.18)	
Memory Disease (%)					0.00 1
	No	4,184 (98.59)	3,555 (98.86)	629 (97.07)	
	Yes	60 (1.41)	41 (1.14)	19 (2.93)	
Arthritis (%)					<0.0 01
	No	2,542 (59.90)	2,208 (61.40)	334 (51.54)	
	Yes	1,702 (40.10)	1,388 (38.60)	314 (48.46)	
Menopause (%)					<0.0 01
	No	1,015 (23.92)	897 (24.94)	118 (18.21)	
	Not applicable	1,723 (40.60)	1,471 (40.91)	252 (38.89)	
	Yes	1,506 (35.49)	1,228 (34.15)	278 (42.90)	
Prostate Disease (%)					0.00 2
	No	1,535 (36.17)	1,326 (36.87)	209 (32.25)	
	Not applicable	2,521 (59.40)	2,125 (59.09)	396 (61.11)	
	Yes	188 (4.43)	145 (4.03)	43 (6.64)	
Hearing Ability (%)					<0.0 01

		Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>value</i>
	Poor	442 (10.41)	337 (9.37)	105 (16.20)	
	Fair	2,222 (52.36)	1,869 (51.97)	353 (54.48)	
	Good	750 (17.67)	650 (18.08)	100 (15.43)	
	Very good	736 (17.34)	655 (18.21)	81 (12.50)	
	Excellent	94 (2.21)	85 (2.36)	9 (1.39)	
Visual Impairment					0.00
(%)					2
	No	783 (18.45)	692 (19.24)	91 (14.04)	
	Yes	3,461 (81.55)	2,904 (80.76)	557 (85.96)	
Pain (%)					<0.0
	No	3,145 (74.10)	2,743 (76.28)	402 (62.04)	01
	Yes	1,099 (25.90)	853 (23.72)	246 (37.96)	
Weight Change (%)					<0.0
	Don't Know	207 (4.88)	176 (4.89)	31 (4.78)	01
	No	3,437 (80.98)	2,951 (82.06)	486 (75.00)	
	Yes, first gained and then lost weight	40 (0.94)	31 (0.86)	9 (1.39)	
	Yes, first lost and then gained weight	28 (0.66)	24 (0.67)	4 (0.62)	
	Yes, only gained weight	250 (5.89)	194 (5.39)	56 (8.64)	
	Yes, only lost weight	282 (6.64)	220 (6.12)	62 (9.57)	
Childhood Health					0.08
(%)					3
	Poor	223 (5.25)	177 (4.92)	46 (7.10)	
	Fair	942 (22.20)	815 (22.66)	127 (19.60)	
	Good	786 (18.52)	659 (18.33)	127 (19.60)	
	Very good	1,837 (43.28)	1,563 (43.46)	274 (42.28)	

	Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>value</i>
Excellent	456 (10.74)	382 (10.62)	74 (11.42)	
BMI (median [IQR]), kg/m²	26.32 [25.03, 28.10]	26.25 [25.01, 28.03]	26.66 [25.27, 28.42]	<0.001
WC (median [IQR]), cm	93.30 [88.70, 98.60]	93.00 [88.50, 98.30]	94.80 [89.97, 100.00]	<0.001
Grip_Dominant (%)				0.947
Right	3,754 (88.45)	3,179 (88.40)	575 (88.73)	
Left	303 (7.14)	257 (7.15)	46 (7.10)	
Both	187 (4.41)	160 (4.45)	27 (4.17)	
SBP (median [IQR]), mmHg	127.50 [116.00, 141.50]	126.50 [115.50, 140.50]	132.50 [119.38, 147.00]	<0.001
DBP (median [IQR]), mmHg	76.50 [69.50, 84.00]	76.00 [69.00, 83.50]	78.00 [70.50, 86.00]	<0.001
WBC (median [IQR]), 10⁹/L	5.81 [4.80, 7.12]	5.87 [4.83, 7.17]	5.70 [4.78, 7.02]	0.012
Hemoglobin (median [IQR]), g/dL	13.30 [11.60, 14.60]	13.20 [11.60, 14.60]	13.45 [11.97, 14.70]	0.037
Hematocrit (median [IQR]), %	40.40 [36.00, 44.30]	40.40 [35.77, 44.30]	40.40 [36.90, 44.50]	0.14
MCV (median [IQR]), fL	91.15 [87.20, 94.80]	91.20 [87.35, 94.90]	90.85 [86.70, 94.50]	0.103
PLT (median [IQR]), 10⁹/L	206.00 [163.00, 253.00]	206.00 [163.00, 253.00]	206.00 [162.00, 251.00]	0.717
CRP (median [IQR]), mg/L	1.70 [1.00, 2.70]	1.70 [1.00, 2.70]	1.80 [1.00, 2.90]	0.066
Creatinine (median [IQR]), mg/dL	0.74 [0.64, 0.86]	0.74 [0.64, 0.86]	0.74 [0.64, 0.86]	0.931
BUN (median [IQR]), mg/dL	15.69 [12.89, 18.77]	15.69 [12.89, 18.77]	15.69 [12.89, 18.49]	0.77

	Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> <i>value</i>
UA (median [IQR]), mg/dL	4.90 [4.10, 5.80]	4.90 [4.10, 5.90]	4.90 [4.20, 5.80]	0.865
Cystatin C (median [IQR]), mg/L	0.82 [0.72, 0.94]	0.82 [0.71, 0.93]	0.86 [0.76, 0.96]	<0.001
TC (median [IQR]), mg/dL	193.44 [168.73, 232.05]	193.44 [168.34, 232.53]	194.98 [171.43, 228.09]	0.082
TG (median [IQR]), mg/dL	151.33 [102.65, 251.33]	150.44 [100.88, 247.79]	160.62 [108.63, 276.11]	<0.001
HDL (median [IQR]), mg/dL	47.10 [40.23, 54.05]	47.10 [40.54, 54.05]	46.33 [40.06, 53.67]	0.015
LDL (median [IQR]), mg/dL	109.65 [89.58, 137.07]	109.65 [89.86, 137.45]	109.65 [88.42, 135.23]	0.232
FBG (median [IQR]), mg/dL	100.90 [91.89, 124.32]	100.90 [91.89, 124.32]	100.90 [91.89, 124.32]	0.436
HbA1c (median [IQR]), %	5.90 [5.60, 6.30]	5.90 [5.60, 6.30]	6.00 [5.60, 6.50]	<0.001
House Type (%)				0.526
Modern	3,713 (87.49)	3,140 (87.32)	573 (88.43)	
Temporary	98 (2.31)	81 (2.25)	17 (2.62)	
Traditional	381 (8.98)	328 (9.12)	53 (8.18)	
Other	52 (1.23)	47 (1.31)	5 (0.77)	
Cooking Fuel (%)				0.006
Clean fuel	2,505 (59.02)	2,156 (59.96)	349 (53.86)	
Non-clean fuel	1,669 (39.33)	1,378 (38.32)	291 (44.91)	
Other/Not cooking	70 (1.65)	62 (1.72)	8 (1.23)	
Room Temperature (%)				0.416
Very cold	2 (0.05)	2 (0.06)	0 (0.00)	

	Total (N = 4,244)	Non-CKM (N= 3,596)	CKM (N=648)	<i>P</i> value
Cold	100 (2.36)	82 (2.28)	18 (2.78)	
Bearable	3,601 (84.85)	3,039 (84.51)	562 (86.73)	
Hot	337 (7.94)	298 (8.29)	39 (6.02)	
Very hot	20 (0.47)	17 (0.47)	3 (0.46)	
Not applicable	184 (4.34)	158 (4.39)	26 (4.01)	

CKM, cardiovascular-kidney-metabolic; IQR, interquartile range; MET, metabolic equivalent of task; BMI, body mass index; WC, waist circumference; SBP, systolic blood pressure; DBP, diastolic blood pressure; WBC, white blood cell count; MCV, mean corpuscular volume; PLT, platelet count; CRP, C-reactive protein; BUN, blood urea nitrogen; UA, uric acid; TC, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; FBG, fasting blood glucose; HbA1c, hemoglobin A1c.

Supplementary Table 4. Multicollinearity assessment using generalized variance inflation factors

	GVIF	Df	GVIF^{1/(2*Df)}
Dyslipidemia	1.093460311	1	1.045686526
Heart Disease	1.132668528	1	1.064269011
Hypertension	1.183766131	1	1.08801017
Kidney Disease	1.118907628	1	1.057784301
Pain	1.136808875	1	1.066212397
Depression	1.127196022	1	1.061694882
Health Expectation	1.139924151	4	1.016504942
Weight Change	1.067659226	5	1.006568339
Age	1.090125038	1	1.044090531

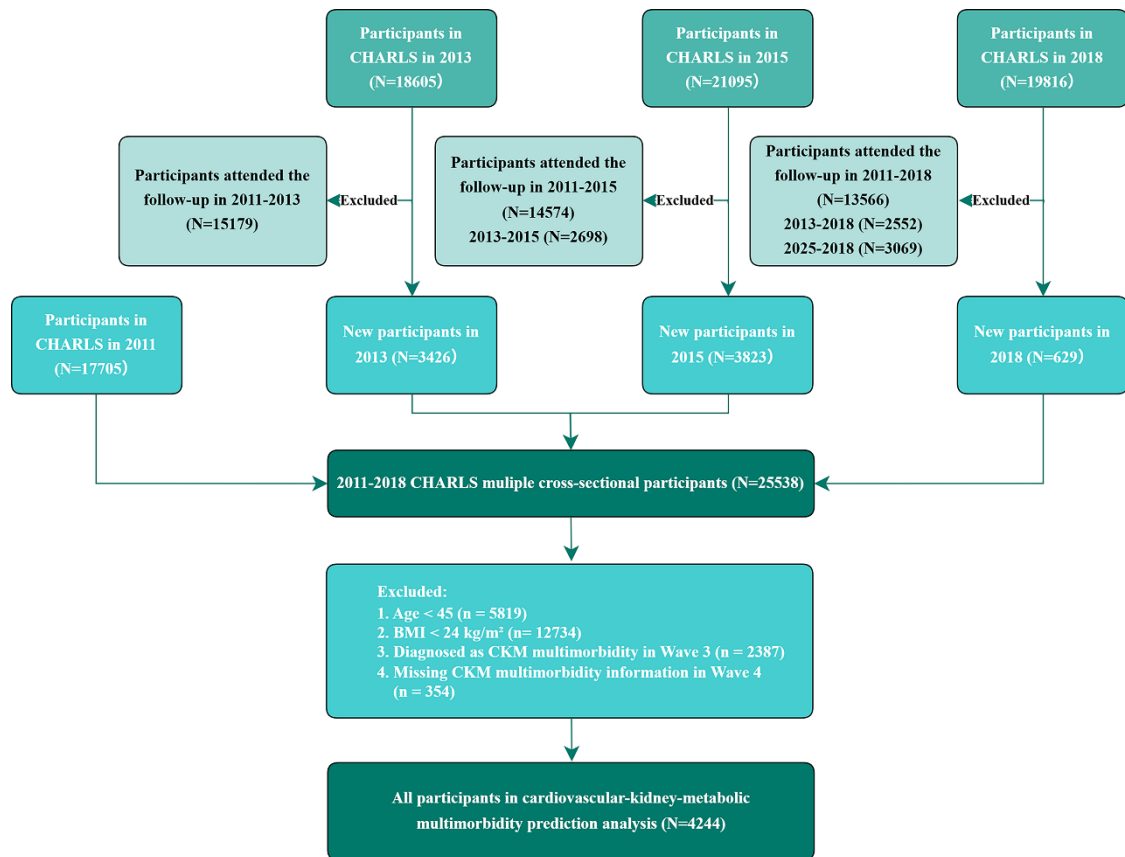
Supplementary Table 5. Classification performance metrics of seven machine learning models in the testing set

Model	Threshold	Sensitivity	Specificity	Accuracy	PPV	NPV	F1
Logistic	0.21	0.72	0.69	0.70	0.30	0.93	0.42
ANN	0.2	0.74	0.68	0.69	0.29	0.94	0.42
Random Forest	0.23	0.69	0.71	0.71	0.31	0.93	0.43
XGBoost	0.22	0.71	0.70	0.70	0.30	0.93	0.42
LightGBM	0.21	0.72	0.70	0.70	0.30	0.93	0.42
SVM	0.25	0.68	0.72	0.72	0.32	0.92	0.43
Decision Tree	0.19	0.75	0.67	0.68	0.28	0.94	0.41

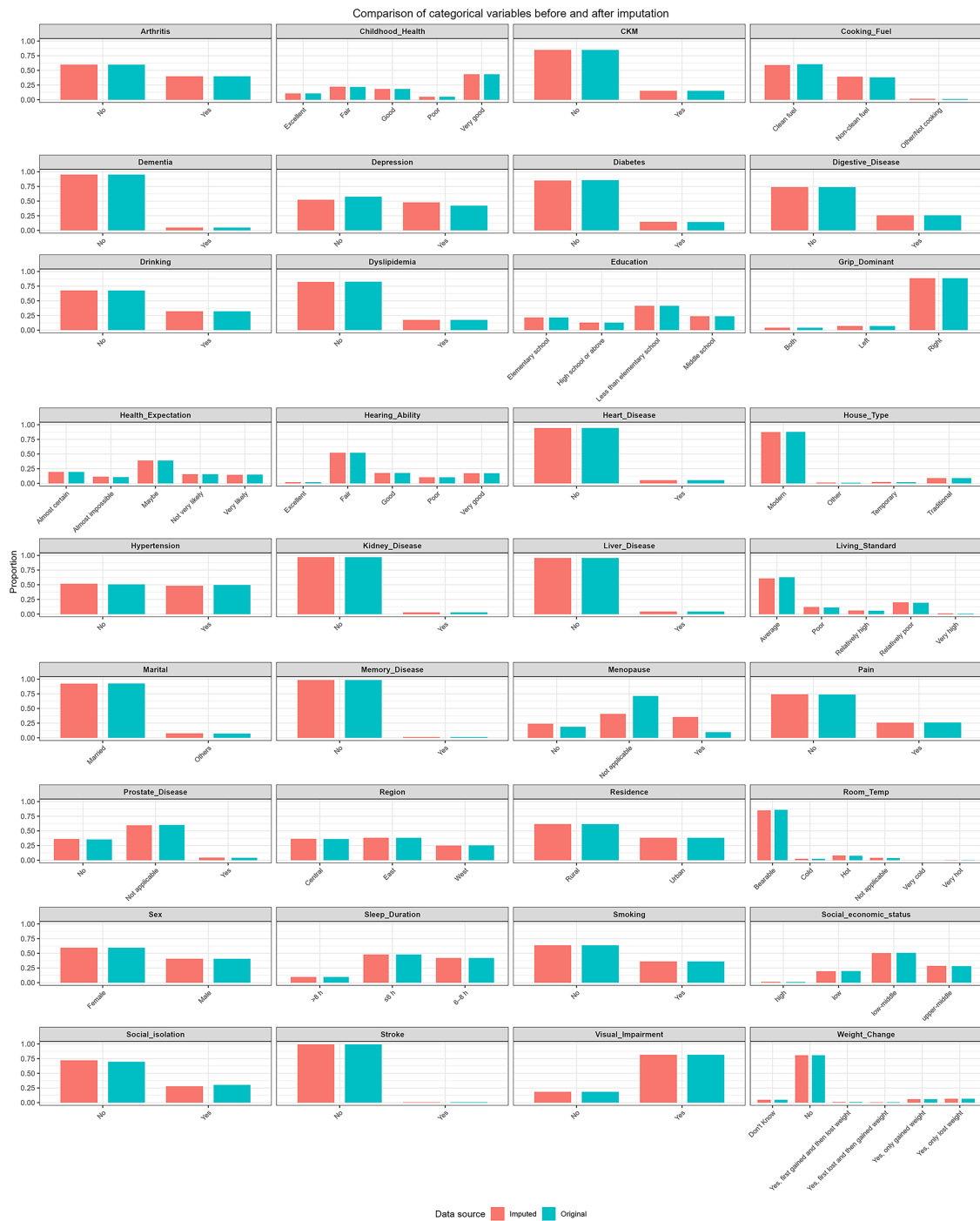
Abbreviations: ANN, artificial neural network; LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machine; DT, decision tree; PPV, positive predictive value; NPV, negative predictive value.

Supplementary Table 6. Pairwise correlation matrix (Spearman's rho) for selected predictors and diabetes

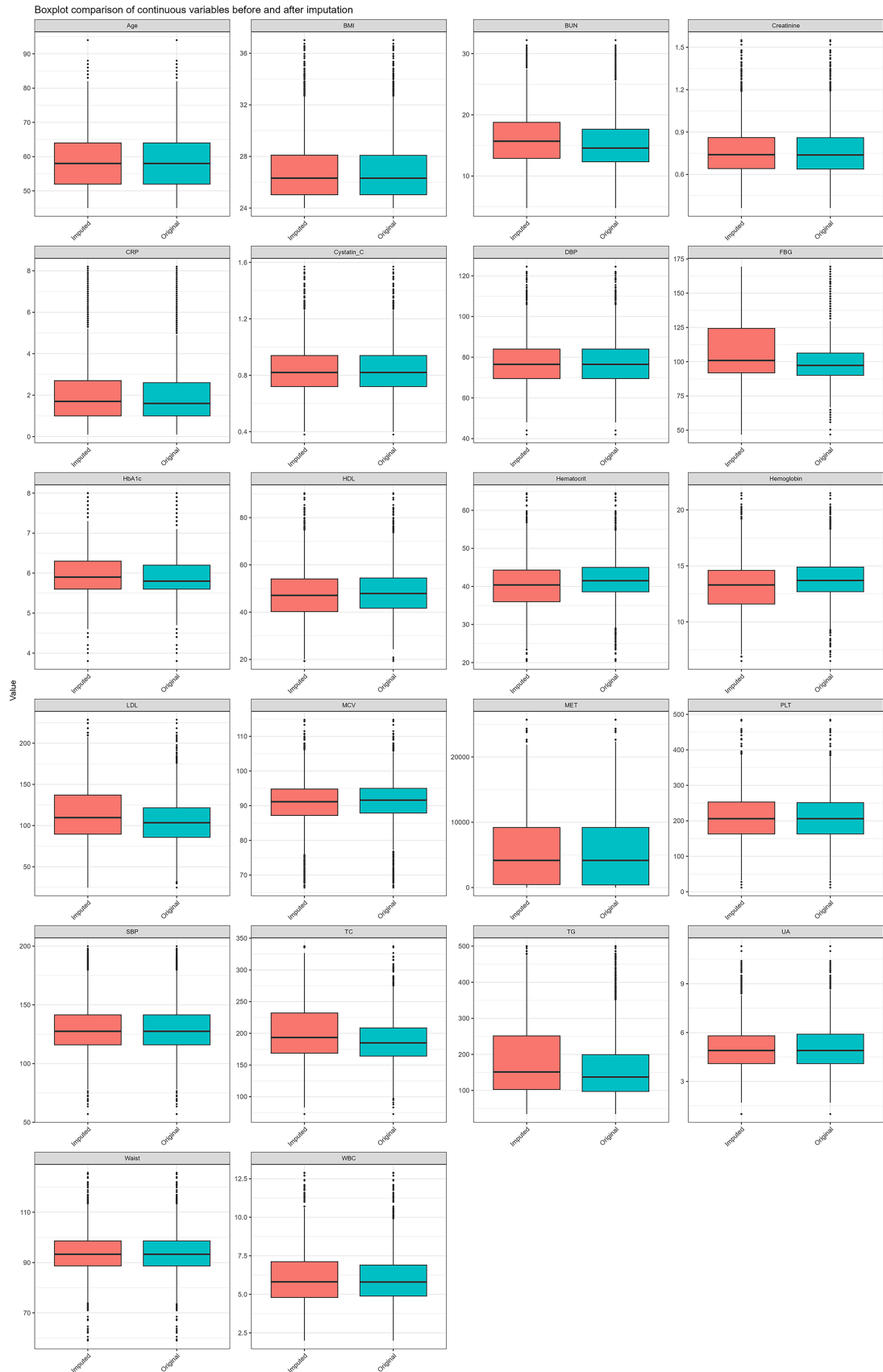
[illegible]



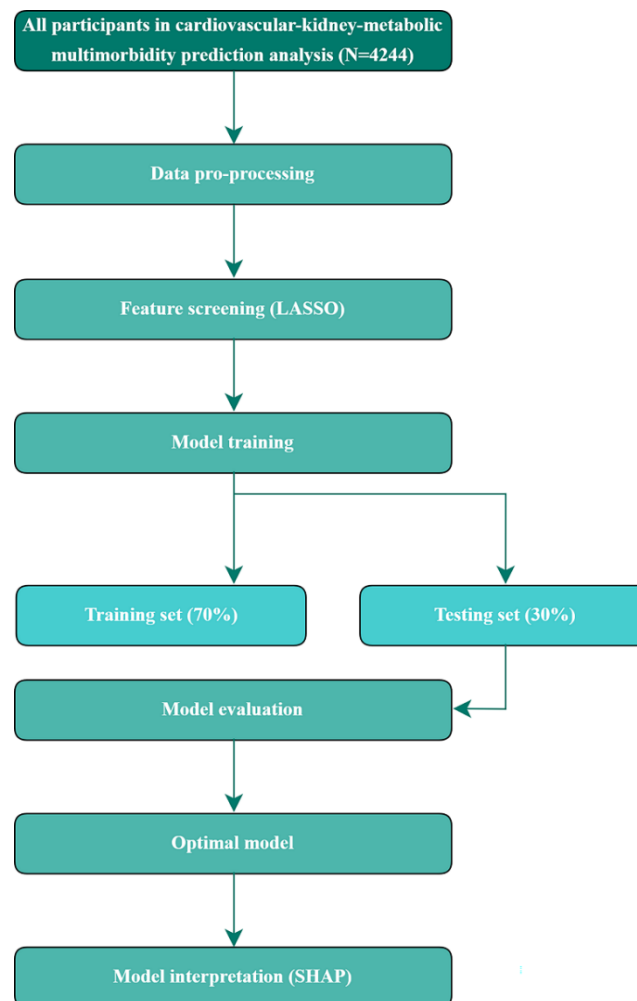
Supplementary Figure 1. Selection process of the study population from the China Health and Retirement Longitudinal Study (CHARLS).



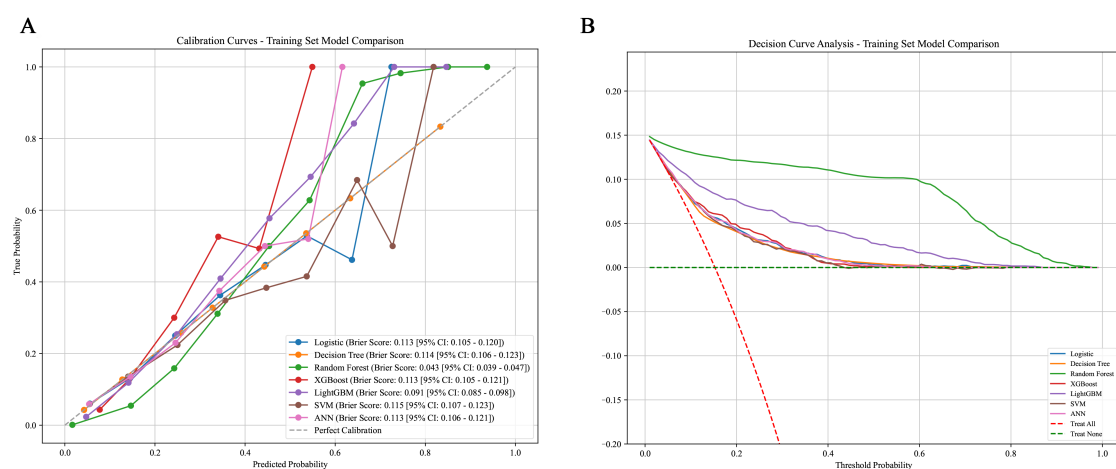
Supplementary Figure 2. Distribution of categorical variables before and after multiple imputation.



Supplementary Figure 3. Boxplots of continuous variables before and after multiple imputation.



Supplementary Figure 4. Workflow of model development.



Supplementary Figure 5. Performance evaluation of seven ML models in the training set. (A) Calibration curves of the seven models, with the Brier score indicated for each model; (B) DCA of the seven models, demonstrating the net clinical benefit across different threshold probabilities. SVM, Support Vector Machine; ANN, Artificial Neural Network.