Acta Cardiol Sin 2023;39:901–912
Original Article
doi: 10.6515/ACS.202311_39(6).20230528A

Artificial Intelligence_AI

Machine Learning Models for ASCVD Risk Prediction in an Asian Population — How to Validate the Model is Important

Yu-Chung Hsiao,[1] Chen-Yuan Kuo,[2] Fang-Ju Lin,[3,4] Yen-Wen Wu,[5,6,7] Tsung-Hsien Lin,[8,9] Hung-I Yeh,[10] Jaw-Wen Chen[11] and Chau-Chung Wu[1,12]

Introduction: Atherosclerotic cardiovascular disease (ASCVD) is prevalent worldwide including Taiwan, however widely accepted tools to assess the risk of ASCVD are lacking in Taiwan. Machine learning models are potentially useful for risk evaluation. In this study we used two cohorts to test the feasibility of machine learning with transfer learning for developing an ASCVD risk prediction model in Taiwan.

Methods: Two multi-center observational registry cohorts, T-SPARCLE and T-PPARCLE were used in this study. The variables selected were based on European, U.S. and Asian guidelines. Both registries recorded the ASCVD outcomes of the patients. Ten-fold validation and temporal validation methods were used to evaluate the performance of the binary classification analysis [prediction of major adverse cardiovascular (CV) events in one year]. Time-to-event analyses were also performed.

Results: In the binary classification analysis, eXtreme Gradient Boosting (XGBoost) and random forest had the best performance, with areas under the receiver operating characteristic curve (AUC-ROC) of 0.72 (0.68-0.76) and 0.73 (0.69-0.77), respectively, although it was not significantly better than other models. Temporal validation was also performed, and the data showed significant differences in the distribution of various features and event rate. The AUC-ROC of XGBoost dropped to 0.66 (0.59-0.73), while that of random forest dropped to 0.69 (0.62-0.76) in the temporal validation method, and the performance also became numerically worse than that of the logistic regression model. In the time-to-event analysis, most models had a concordance index of around 0.70.

Conclusions: Machine learning models with appropriate transfer learning may be a useful tool for the development of CV risk prediction models and may help improve patient care in the future.

Key Words: Atherosclerotic cardiovascular disease • Machine learning • Risk prediction model

Received: September 12, 2022 Accepted: May 28, 2023

[1]Department of Internal Medicine, National Taiwan University Hospital; [2]Center for Healthy Longevity and Aging Sciences, National Yang Ming Chiao Tung University; [3]Graduate Institute of Clinical Pharmacy & School of Pharmacy, College of Medicine, National Taiwan University; [4]Department of Pharmacy, National Taiwan University Hospital, Taipei; [5]Division of Cardiology, Cardiovascular Medical Center, Far Eastern Memorial Hospital, New Taipei City; [6]School of Medicine, National Yang Ming Chiao Tung University, School of Medicine, Taipei; [7]Graduate Institute of Medicine, Yuan Ze University, Taoyuan; [8]Division of Cardiology, Department of Internal Medicine, Kaohsiung Medical University Hospital; [9]Faculty of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung; [10]MacKay Memorial Hospital, MacKay Medical College; [11]Department of Medical Research and Education, Taipei Veterans General Hospital; [12]Graduate Institute of Medical Education & Bioethics, College of Medicine, National Taiwan University, Taipei, Taiwan.

Corresponding author: Dr. Chau-Chung Wu, Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Graduate Institute of Medical Education & Bioethics, College of Medicine, National Taiwan University, No. 7, Chung-Shan South Road, Taipei 10022, Taiwan. Tel: 886-2-2312-3456 ext. 88560; Fax: 886-2- 2321-7485; E-mail: chauchungwu@ntu.edu.tw

Abbreviations

ACEi	Angiotensin-converting enzyme inhibitor
ARB	Angiotensin receptor blocker
ASCVD	Atherosclerotic cardiovascular disease
AUC	Area under the receiver operating characteristic curve
BMI	Body mass index
CAD	Coronary artery disease
CI	confidence interval
CRP	C-reactive protein
CV	Cardiovascular
DM	Diabetes mellitus
eGFR	Estimated glomerular filtration rate
HDL-C	High-density lipoprotein cholesterol
LDL-C	Low-density lipoprotein cholesterol
MACE	Major adverse cardiovascular event
MLP	Multilayer perceptron
PAOD	Peripheral arterial occlusive disease
TG	Triglyceride
XGBoost	eXtreme Gradient Boosting

INTRODUCTION

Atherosclerotic cardiovascular disease (ASCVD) is prevalent worldwide including Taiwan. Previous studies have shown that there are epidemiological and drug response differences between Asian and Western patients. Taiwan has an easily accessible health insurance system that may lead to differences in cardiovascular (CV) disease outcomes from other countries.[1,2] ASCVD risk scoring is important for risk reduction strategies. U.S., European and Asian guidelines recommend the management of ASCVD and metabolic risk factors based on CV risk scoring.[3-7] However, widely accepted tools to assess the risk of ASCVD are lacking in Taiwan.

Machine learning has been extensively used for developing prognostic models in healthcare. It has the advantage of analyzing complex data and relies less on ever-changing medical knowledge. However, machine learning models are often criticized as being difficult to interpret and lacking reproducibility.[8] Transferring knowledge from other cohorts or models and adjusting it according to another cohort, a type of "transfer learning", is increasingly being used in medical classification problems.[9,10] In a broader definition, this also includes referring the concept and variables of other models and then building a new model with our own data.

The purpose of this study was to test the feasibility of machine learning with transfer learning for evaluating ASCVD risk in Taiwan. We used both traditional statistical analysis and machine learning to develop risk prediction models with two Taiwanese cohorts, and then validated the models in various ways.

METHOD

Participants' demographics

The study population constituted the T-SPARCLE and T-PPARCLE cohorts, which enrolled patients from 16 medical centers in Taiwan. The cohorts included men and women aged > 18 years who met the following criteria (Figure 1): (a) with evidence of ASCVD, including (1) coronary artery disease (CAD), evidenced by cardiac catheterization examination, having a history of myocardial infarction, or with angina showing ischemic electrocardiogram changes or positive response to stress test); (2) cerebral vascular disease, cerebral infarction, intracerebral (excluding intracerebral hemorrhage associated with other diseases); (3) transient ischemic attack with carotid artery ultrasound confirming atheromatous changes with more than 70% blockage; or (4) peripheral atherosclerosis (symptoms of ischemia confirmed by Doppler ultrasound or angiography); and (b) with no evidence of ASCVD, but with at least one of the following risk factors: diabetes mellitus (DM), dyslipidemia, hypertension, smoking, older age (men > 45 years old, women > 55 years old), family history of premature CAD (men < 55 years old, women < 65 years old), and obesity (waist circumference: men > 90 cm, women > 80 cm). The patients were defined as having dyslipidemia if one of the following criteria were met: total cholesterol > 200 mg/dL; low-density lipoprotein cholesterol (LDL-C) > 130 mg/dL; triglyceride (TG) > 200 mg/dL; men with high-density lipoprotein cholesterol (HDL-C) < 40 mg/dL or women with HDL-C < 50 mg/dL, or receiving lipid-lowering therapy.

The exclusion criteria were as follows: patients with neuro-cognitive or psychiatric conditions, end-stage renal disease on dialysis, serious heart disease with functional class III or IV heart failure, life expectancy shorter than 6

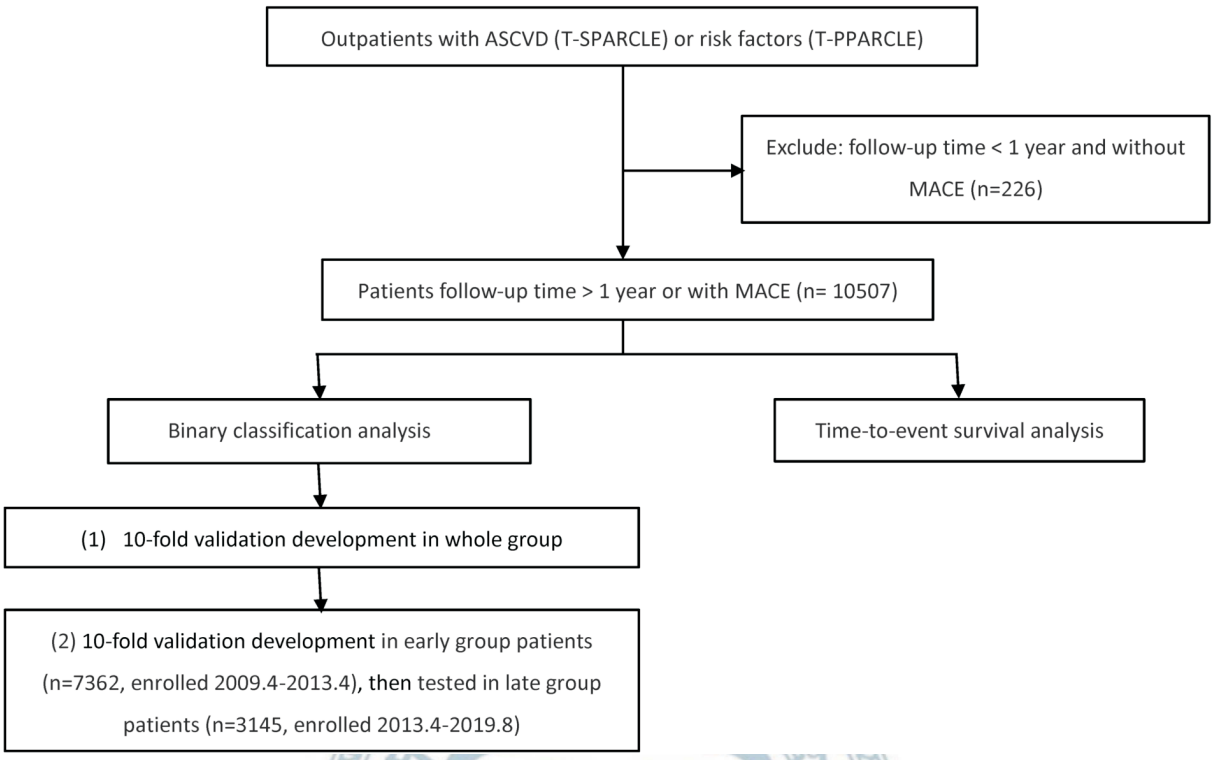


Figure 1. *Data processing and model flow. ASCVD, atherosclerotic cardiovascular disease; MACE, major adverse cardiovascular event.*

months, treatment with immunosuppressive agents. Patients with recent acute stroke, acute myocardial infarction, and those who underwent coronary revascularization within 3 months were also excluded. The enrolled patients were followed yearly, and those with a follow-up time < 1 year but without a major adverse cardiovascular event (MACE) were also excluded. Clinical outcomes, adverse events, laboratory data and medication use were recorded at enrolment and each follow-up. Smoking status, physical activity, and other relevant clinical information were also recorded.

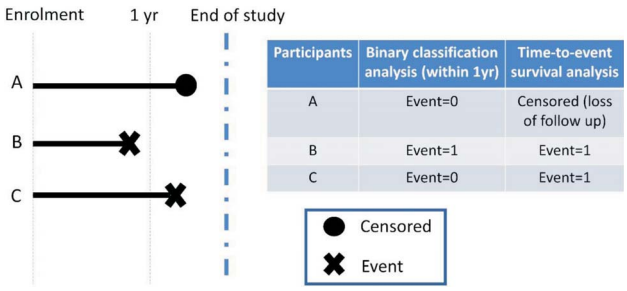
Machine learning analytic framework

In this study, 22 clinical variables were recorded, including age, sex, history of ASCVD, revascularization procedure (performed not for acute coronary syndrome), peripheral arterial occlusive disease (PAOD), stroke, smoking status, systolic blood pressure, congestive heart failure, DM, renal function (estimated glomerular filtration rate), body mass index (BMI), lipid profiles (including LDL-C, non-HDL-C, HDL-C, TG), and medication use [statins, beta-blockers, angiotensin-converting enzyme inhibitors (ACEis)/angiotensin receptor blockers (ARBs), antiplatelets]. The variables were based on European, U.S. and Asian guidelines and studies considered to be important for the risk stratification of ASCVD or management.[4,5,11,12] We proposed two approaches with machine learning and deep learning framework to evaluate the clinical outcomes. First, the patients were defined to have events if they had a MACE within 1 year; otherwise, they were defined as having no events. Binary classification with four different algorithms [logistic regression, eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), and random forest] were constructed to classify the patients into those with and without MACEs in 1 year. Second, to predict MACEs during follow-up, two survival analyses (Cox proportional hazards regression and DeepHit) were conducted (Figure 2).

Binary classification

Logistic regression was performed for the linear analysis in the binary classification method. Other non-linear models were as follows:

XGBoost: a tree-based, efficient, non-linear model that is capable of handling missing data. It is an ensemble method that uses gradient boosting for optimization. It



Participants	Binary classification analysis (within 1yr)	Time-to-event survival analysis
A	Event=0	Censored (loss of follow up)
B	Event=1	Event=1
C	Event=0	Event=1

Figure 2. *Example of binary classification and time-to-event survival analysis.*

has been shown in prior studies to yield state of the art performance.[13]

MLP: a neural network-based model that has one extra multiple hidden layer between the input and output layer. Neurons constitute each hidden layer, and each layer is connected with weight coefficients which are updated during the training process. The loss function is cross-entropy.[14]

Random forest: an estimator consisting of a combination of tree classifiers where each classifier is generated using different random subsamples of input. It is also an ensemble method.[15]

In each of these models, ten-fold cross-validation was performed and the average performance score was presented (Figure 3). Feature importance was displayed using a permutation method after repeating 1000 times.[16] In an additional temporal validation method, 70% of the data were used for training (model development) and 30% were used for testing according to the time of enrollment. The early group was used for training, and the late group was used for testing. Ten-fold validation and model development were performed in the training set (n = 7362), and then validated in the test set (n = 3145) (Figure 4a).

Figure 3. *Process of 10-fold validation.*

Time-to-event survival analysis

In the time-to-event analysis, data were randomly split into 70% for training and 30% for testing. In the training set, 70% was used for model development, and the remaining 30% was used for model tuning and selection (Figure 4b).

Cox proportional hazards regression was performed for linear analysis of the time-to-event survival analysis. The loss function was Cox's partial likelihood. Other non-linear models were as follows:

DeepHit: a deep neural-network based model that can learn survival time. The model does not rely on parametric assumptions of survival. The loss function is a combination of a negative log-likelihood and a ranking loss.[17]

Gradient boosted survival model: an ensemble tree-

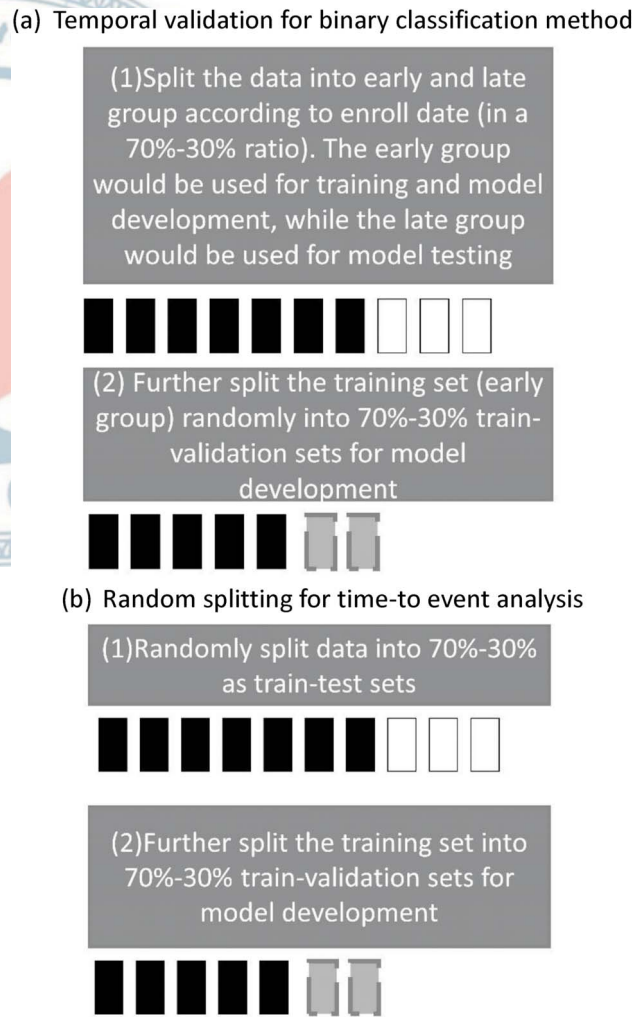


Figure 4. *Process of 70%-30% train-test splitting.*

based method that uses a gradient boost method to optimize loss function in time-to-event analysis. The loss function used here is also Cox's partial likelihood.[18]

SMART score

SMART score, a Dutch scoring system, was used to compare with the models developed in this study. Model A in the original study was chosen, and variables included age, sex, smoking, systolic blood pressure, DM, CAD, cerebrovascular disease, PAOD, HDL-C, total cholesterol, and eGFR; while C-reactive protein (CRP), abdominal aneurysm and timing of vascular disease were omitted in the study due to a lack of data in our cohort.[19]

For the binary outcome analytic model, performance was measured according to area under the receiver operating characteristic (ROC) curve (AUC). The average precision score was also calculated, which is the AUC of precision (also called positive predictive value) recall (also called sensitivity) curve. For the time-to-event model, the performance was measured using the concordance index.

All analyses were performed using SAS 9.4 software (SAS Institute Inc., Cary, NC) and in Python (version 3.3.4) using open-source libraries (Matplotlib [version 3.6.9], XGBoost [version 1.5.1]).

Statistical analysis

Categorical variables are presented as percentage, and continuous or discrete variables are presented as mean ± standard deviation. The chi-square test was used to compare proportions, and the Student's t-test was used to compare differences in continuous variables between groups.

A Cox regression model was used to estimate hazard ratios and 95% confidence intervals (CIs) for MACE outcomes among CV risk groups, and evaluate the residual risk related to lipid profiles. Missing data were imputed with their means. The CIs of AUCs were calculated assuming exponential distribution of positive and negative cases,[20] while that of the concordance index was obtained from bootstrapping of the testing data.

RESULTS

After splitting the data according to the time of enrollment, those enrolled by April 8, 2013 were classified into the early group (n = 7362), while those enrolled after April 8, 2013 were classified into the late group (n = 3145). The baseline characteristics are listed in Table 1. Compared to the early group, the late group had higher event rates, more male patients, older age, higher BMI, higher eGFR, lower rate of stroke, lower rate of myocardial infarction, more ASCVD, more coronary revascularizations (not for ACS), lower systolic blood pressure, lower HDL-C, higher TG, and less frequent use of ARBs/ACEIs. Compared to those without MACEs within 1 year, those with MACEs within 1 year were older, had lower eGFR, and were more likely to have DM, ASCVD, stroke, PAOD, myocardial infarction, heart failure and to use antiplatelets. Pearson correlation coefficients between the variables are provided in Supplementary Table 1.

For the binary classification methods, when the whole group was used for ten-fold cross-validation, XGBoost and random forest had numerically higher ROC-AUC values [0.72 (0.68-0.76), 0.73 (0.69-0.77), respectively] and higher average precision [0.18 (0.13-0.23), 0.17 (0.12-0.22), respectively]. All models had substantial overlap in CI (Figure 5). The SMART score had an ROC-AUC of 0.70 (0.66-0.74), and average precision of 0.04 (0.01-0.07). In the logistic regression model, the five most important features were PAOD, DM, TG, heart failure, and ASCVD; those in the XGBoost model were eGFR, age, ASCVD, BMI, and DM; those in the MLP model were age, systolic blood pressure, revascularization, DM, and ARB/ACEi use; and those in the random forest were age, eGFR, ASCVD, BMI, and HDL.

For the temporal validation methods, XGBoost had the best performance in the training set (early group), with an ROC-AUC of 0.73 (0.68-0.78), while the performance dropped to 0.66 (0.59-0.73) in the test set (late group). Similarly, random forest had a performance of 0.72 (0.66-0.78) in the training set, and 0.69 (0.62-0.76) in the test set. In contrast, both the logistic regression and MLP models had slightly better performance for the test set than that for the training set (Table 2).

For the time-to-event analyses, the Cox regression model had a concordance index of 0.69 in the training set, and 0.70 in the test set. The DeepHit model had a concordance index of 0.65 in both the training and test sets. The gradient boost survival model had a concordance index of 0.70 in both the training and test sets.

Table 1. Baseline characteristics of the patients

	Early group (N = 7362, by 2013/4/8)	Late group (N = 3145, after 2013/4/8)	p value	No MACE within 1 year (N = 10411)	MACE within 1 year (N = 96)	p value
Age	64.6 ± 11.9	63.3 ± 12.0	< 0.001	64.2 ± 11	69.4 ± 12.7	< 0.001
Male sex (%)	4538 (61.6)	2126 (67.6)	< 0.001	6602 (63.4)	62 (64.6)	0.813
ASCVD (%)	3995 (54.3)	1739 (55.3)	0.332	5656 (54.3)	78 (81.3)	< 0.001
Revascularization (not for ACS)	80 (1.1)	349 (11.1)	< 0.001	423 (4.1)	6 (6.3)	0.281
Unstable angina	71 (1.0)	106 (3.4)	< 0.001	174 (1.7)	3 (3.1)	0.270
Stroke	865 (11.7)	211 (6.7)	< 0.001	1056 (10.1)	20 (20.8)	0.001
PAOD	108 (1.5)	28 (0.9)	0.020	130 (1.2)	6 (6.3)	< 0.001
Myocardial infarction	2890 (39.3)	1047 (33.3)	< 0.001	3881 (36.9)	56 (58.3)	< 0.001
Heart failure	552 (7.5)	362 (11.5)	< 0.001	895 (8.6)	19 (19.8)	< 0.001
BMI	26.2 ± 3.9	25.4 ± 4.5	< 0.001	26.3 ± 4.1	26.7 ± 4.8	0.348
eGFR	75.9 ± 23.6	82.9 ± 27.3	< 0.001	78.1 ± 25.0	66.4 ± 26.3	< 0.001
DM (%)	3138 (42.6)	1331 (42.30)	0.773	4407 (42.3)	62 (64.6)	< 0.001
Smoking (%)	1605 (21.8)	711 (22.6)	0.361	2291 (22.0)	25 (26.0)	0.361
SBP (mmHg)	133.8 ± 17.7	132.3 ± 18.3	< 0.001	133.3 ± 17.9	136.2 ± 20.9	0.119
HDL-C	47.8 ± 13.5	45.0 ± 13.5	< 0.001	47.0 ± 13.5	43.2 ± 13.1	0.006
TG	141.0 ± 94.0	151.2 ± 112.4	< 0.001	143.7 ± 98.4	185.8 ± 207.7	0.050
LDL-C	103.7 ± 33.0	103.8 ± 35.1	0.955	103.8 ± 33.6	99.3 ± 35.7	0.222
Non-HDL-C	131.3 ± 35.7	132.8 ± 37.8	0.048	131.3 ± 36.3	134.3 ± 45.5	0.583
Statin	4402 (59.8)	1882 (59.8)	0.964	6230 (59.8)	54 ± 56.3	0.475
Antiplatelets	4522 (61.4)	1924 (61.2)	0.812	6373 (61.2)	73 (76)	0.003
ARB/ACEI	4240 (57.6)	1739 (55.3)	0.030	5924 (56.9)	55 (57.3)	0.900
Beta blockers	3652 (49.6)	1878 (59.7)	< 0.001	5483 (52.7)	47 (49.0)	0.469
MACE within 1 yr	54 (0.7)	42 (1.7)	0.003			

ACS, acute coronary syndrome; ARB/ACEi, angiotensin receptor blocker/angiotensin-converting enzyme inhibitor; ASCVD, atherosclerotic cardiovascular disease; BMI, body mass index; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MACE, major adverse cardiac events; Non-HDL-C, non high-density lipoprotein cholesterol; PAOD, peripheral arterial occlusive disease; SBP, systolic blood pressure; TG, triglyceride.

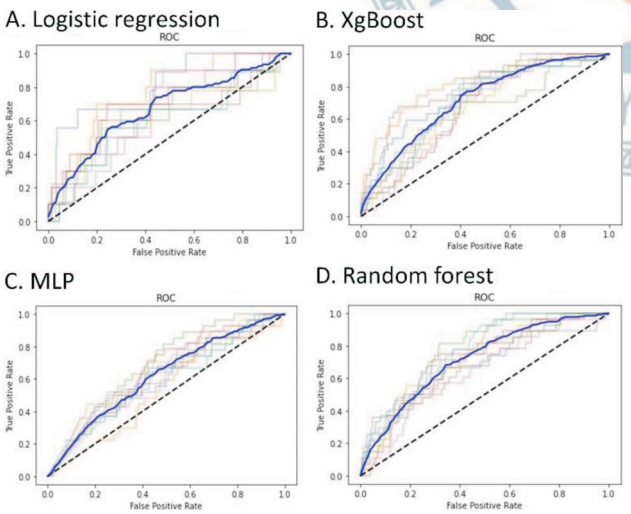


Figure 5. *Comparing performance between models. In each receiver operating characteristic (ROC)-area under the curve (AUC) figure, color lines represent performances of 10 different validation process, while the central blue line is the performance of the final model. MLP, multilayer perceptron; XGBoost, eXtreme Gradient Boosting.*

The SMART score was validated on the whole data set, and had a concordance index of 0.69 (Table 3).

DISCUSSION

In this study, we demonstrated that machine learning is a potentially useful tool for developing a cardiovascular risk scoring system in Taiwan. We also presented the results of various validation methods to ensure the reproducibility of the models with two clinical cohorts (Central Illustration).

Different approaches to predict cardiovascular outcomes yielded similar performance in this study. Both binary classification and time-to-event analyses had AUCs or concordance indices of around 0.7. The performance was similar to that of the traditional Framingham score endorsed by the AHA but numerically inferior to that of

Table 2-1. Binary classification analysis by method (1): ten-fold validation development in whole group

Model	SMART score*	Logistic	XGBoost	MLP (neural network)	Random Forest
ROC-AUC (n = 10507)	0.70 (0.66-0.74)	0.67 (0.62-0.71)	0.72 (0.68-0.76)	0.62 (0.57-0.67)	0.73 (0.69-0.77)
Average precision (n = 10507)	0.04 (0.01-0.07)	0.07 (0.03-0.11)	0.18 (0.13-0.23)	0.02 (0.00-0.04)	0.17 (0.12-0.22)
Most important features	Age, eGFR, stroke, PAOD	eGFR, ASCVD, DM, BMI, antiplatelet	eGFR, systolic blood pressure, BMI, age, LDL-C	eGFR, Non-HDL-C, age, LDL-C, ASCVD	Age, eGFR, ASCVD, BMI, HDL

* Not cross validated.
ASCVD, atherosclerotic cardiovascular disease; BMI, body mass index; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; LDL-C, low-density lipoprotein cholesterol; MLP, multilayer perceptron; PAOD, peripheral arterial occlusive disease. ROC-AUC, area under the curve of the receiver operating characteristic curve; XGBoost, eXtreme Gradient Boosting.

Table 2-2. Binary classification analysis by method (2): temporal validation

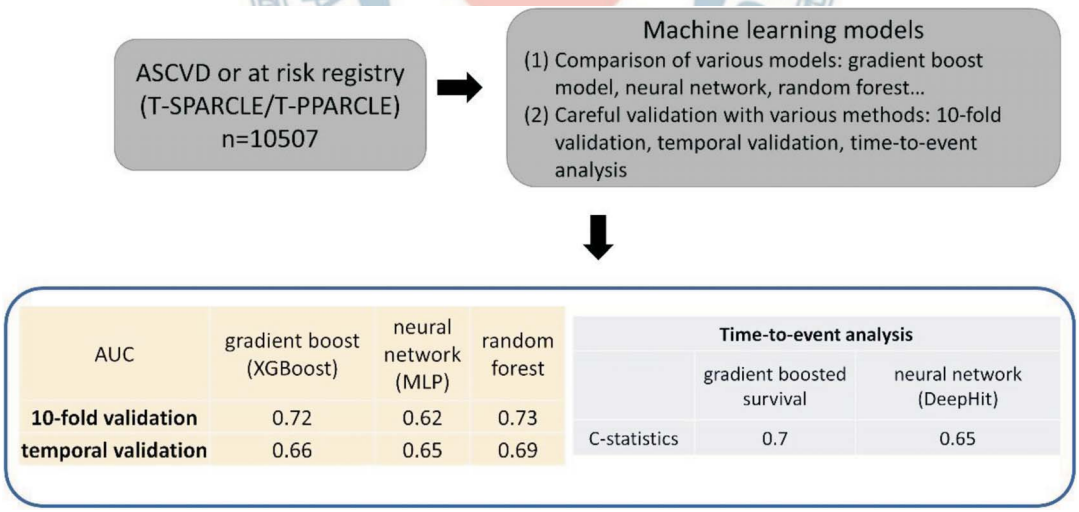
Model	SMART	Logistic	XGBoost	MLP (neural network)	Random Forest
Early group as train set, ROC-AUC (ten-fold validation, n = 7362)	0.70 (0.66-0.74)*	0.63 (0.56-0.70)	0.73 (0.68-0.78)	0.60 (0.53-0.67)	0.72 (0.66-0.78)
Late group as test set, ROC-AUC (n = 3145)	0.70 (0.64-0.76)	0.68 (0.61-0.75)	0.66 (0.59-0.73)	0.65 (0.58-0.72)	0.69 (0.62-0.76)
Test average precision (n = 3145)	0.04 (0-0.08)	0.06 (0.01-0.11)	0.04 (0-0.08)	0.04 (0-0.08)	0.06 (0.01-0.11)

* Not cross validated. MLP, multilayer perceptron; ROC-AUC, area under the curve of the receiver operating characteristic curve; XGBoost, eXtreme Gradient Boosting.

Table 3. Time-to-event analysis

Model/C-index	SMART score	Cox proportional hazards regression	DeepHit	Gradient boosted survival
Train*	0.69 (0.67-0.72)	0.69	0.65	0.70
Test		0.70 (0.64-0.73)	0.65 (0.60-0.69)	0.70 (0.60-0.71)

* 30% of the trained set was used as validation.



Central Illustration. *Strategies for rigorous validation of machine learning models. ASCVD, atherosclerotic cardiovascular disease; AUC, area under the curve; MLP, multilayer perceptron; XGBoost, eXtreme Gradient Boosting.*

SCORE2 endorsed by the European Society of Cardiology, which have been validated in other corhorts.[21-23] Both SCORE and Framingham score were shown to be useful for risk discrimination in Asians in a prior study,[24] however they are both used to evaluate the risk of patients without ASCVD. In contrast, the SMART score was devel-

oped for use in patients with ASCVD. To the best of our knowledge, this is the first study to validate the SMART score in an Asian cohort.

Even with the much smaller sample size, this study had similar performance to that of studies from the UK and Korea, in which machine learning models were also used (Table 4).[25,26] Of note, we used fewer parameters in the current study. All of the models used in the UK study were also tested in our cohort. In the Korean study, the researchers used more variables, including history of atrial fibrillation, rheumatoid arthritis, systemic lupus erythematous, migraine, and the use of corticosteroids and atypical antipsychotic medications. However, the Korean study did not use variables of heart failure, stroke or specified types of ASCVD, which were used in our study. Our study has the strength of using additional validation methods in order to minimize overestimation of model performance. Predicting events in the near future may be achievable.[27] For example, other studies of using machine learning to predict in-hospital mortality after myocardial infarction have reported AUCs of 0.8 to 0.9.[28-30] All of the variables used in this study were recorded at enrolment. The low average precision score is common in longer risk prediction models, and is related to a low event rate.[31] Future studies with a larger cohort and longer follow-up are likely to increase the accuracy of prediction and stability of the performance.

Ideally, external validation should be used to confirm the performance of machine learning models.[32,33] Temporal validation, splitting the data based on time of enrollment, is considered to be analogous to external validation and may make prediction models more generalizable.[34] The results of temporal validation in our study suggest that further model adjustment may be required before applying the risk prediction model to new cohorts. In the temporal validation approach, there were some significant differences in baseline risk factors between the early and late groups in this study. The event rates were also different, and the patients enrolled later had a higher MACE rate. Such differences may also be encountered in real-world scenarios. Careful analysis of the validation dataset and adjusting it accordingly may be required before application.[32,35] This concept is essential if attempting to use "transfer learning", that is, applying models from Western cohorts to Taiwanese cohorts. On the other hand, machine learning models may still be subject to overfitting despite careful adjustments. Even though cross-validation is a widely used method for model validation,[36] the XGBoost model in our study did not generalize the results very well in temporal validation. The performance of the SMART score, a European scoring system, seemed to have a more stable performance. This implies that external validation or temporal validation should be performed to confirm the robustness of machine learning models.

Among the many machine learning models available for binary outcome classification, XGBoost often yields superior accuracy compared to others.[37] One survey showed that random forest may also have similar performance.[38] MLP classifier is also a commonly used model that can exhibit excellent performance, but the hyperparameters are difficult to tune.[39] Random forest models are gaining popularity as they are faster to train and have good accuracy. Prior studies have demonstrated that machine learning did not consistently have advantages over logistic regression in different cohorts.[40] Our studies have shown considerable overlap of the important features of different models. If the relationships between the features and outcomes were mostly linear, the non-linear machine learning models did not neces-

Table 4. Comparison of performance of various machine learning models in different cohorts

Cohort	Baseline risk	Sample size	Variables used	Best algorithm	AUC	Validation
UK biobank[28]	No CVD	423,604	473	Ensemble of XgBoost, neural network, random forest	0.77	Internal validation: ten-fold validation
Korean[27]	No CVD, statin naive	222,998	16	Neural network	0.75	Internal validation: Random split 70-30%
Taiwan (this work)	With or without CVD	10,507	22	XgBoost, random forest	0.73	Internal validation: ten-fold validation, temporal validation, time-to-event analysis

AUC, area under the receiver operating characteristic curve; CVD, cardiovascular disease.

sarily have superior performance to the logistic regression model. However, in the present study we demonstrated that using a machine learning model is much easier for beginners in the research field to develop ASCVD risk scoring systems with at least modest performance comparable to that of traditional methods which may require more mature medical and complicated statistical knowledge. With a larger sample size, more clinical variables, and further machine learning, the developed models may evolve and perform better for this population. This may in turn lead to the development of good ASCVD risk scoring systems in Taiwan.

Having serial follow-up data of the variables as input may enhance model performance. One study using updated follow-up data of a cancer cohort achieved an AUC of 0.79 with a sample size of only 585.[41] In addition, a previous meta-analysis reported that machine learning for predicting chronic coronary artery disease based on imaging had an AUC ranging from 0.8 to 0.9.[42] Some algorithms rely on advanced biomarkers (e.g. tissue necrosis factor-α soluble receptor, interleukin-2 soluble receptor) and imaging (e.g. intravascular ultrasound, cardiac magnetic resonance imaging, single photon emission computed tomography), which are not used by the aforementioned conventional tools.[43,44] It is likely that incorporating genetic profile, biomarkers such as CRP and lipoprotein (a)[45] and imaging such as CT-angiography and echocardiography will enhance the prediction accuracy.[46] In addition, novel agents such as gliflozins and glucagon-like peptide 1 agonists are also likely to alter the course of cardiovascular disease in the future, and newer risk prediction models are needed. Machine learning is likely to aid the interpretation of high dimensional features in the future.

There are several limitations to this study. First, the performances of our models were only modest. Second, we did not prove that each hyperparameter used in the models was already the best. Third, although the performance of XGBoost was shown to be overly optimistic in the training set of the temporal validation method, we did not present solutions to the problems met in the test set. Hence, further tuning of the models may improve performance. With a larger sample size, more clinical variables, and further machine learning, the developed models may evolve and perform better for this population. There are still other problems with machine learning, including ethical issues, data preparation, model robustness, and model explainability, which need to be managed in the future.[47,48] The strength of this study is that we show consistent performance using limited resources.

CONCLUSIONS

In conclusion, machine learning algorithms with transfer learning may be a useful tool for the development of CV risk prediction models and may help to improve patient care in the future. The use of larger cohorts, successful knowledge transfer, incorporating novel parameters and careful model tuning and validation are important for future machine learning application to health care in Taiwan.

ACKNOWLEDGEMENT AND FUNDING

This study is supported by Taiwan Society of Lipids & Atherosclerosis since 2009, Taiwan Ministry of Science and Technology since 2012 (Project code: NRPB-TR11: 100-2325-B-002-075, 102-2325-B-002-097, 104-2325-B-002-036, 105-2325-B-002-035, 106-2321-B-002-029, 107-2321-B-002-035, 108-2321-B-002-043, 109-2321-B-002-039, and 110-2745-B-002 -004), and Taiwan Association of Lipid Educators since 2018.

DECLARATION OF CONFLICT OF INTEREST

All the authors declare no conflict of interest.

REFERENCES

1. Wu TY, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. *London J Prim Care* 2010;3:115-9.
2. Wu CK, Juang JJ, Chiang JY, et al. The Taiwan heart registries: its influence on cardiovascular patient care. *J Am Coll Cardiol* 2018; 71:1273-83.
3. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hyperten-

sion (ESH). *Eur Heart J* 2018;39:3021-104.
4. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk: The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS). *Eur Heart J* 2020;41: 111-88.
5. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Circulation* 2019;139:e1082-143.
6. Unger T, Borghi C, Charchar F, et al. 2020 International Society of Hypertension global hypertension practice guidelines. *Hypertension* 2020;75:1334-57.
7. Li YH, Ueng KC, Jeng JS, et al. 2017 Taiwan lipid guidelines for high risk patients. *J Formos Med Assoc* 2017;116:217-48.
8. Heil BJ, Hoffman MM, Markowetz F, et al. Reproducibility standards for machine learning in the life sciences. *Nat Methods* 2021;18:1132-5.
9. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data* 2016;3:9.
10. Ebbehoj A, Thunbo MØ, Andersen OE, et al. Transfer learning for non-image data in clinical research: a scoping review. *PLoS Digit Health* 2022;1:e0000014.
11. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). *Eur Heart J* 2021;42:3227-337.
12. Honda T, Chen S, Hata J, et al. Development and validation of a risk prediction model for atherosclerotic cardiovascular disease in Japanese adults: The Hisayama Study. *J Atheroscler Thromb* 2022;29:345-61.
13. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System: Association for Computing Machinery, New York, United States; 2016.
14. Svozil D, Kvasnicka V, Pospichal Jí. Introduction to multi-layer feed-forward neural networks. *Chemometr Intell Lab Syst* 1997; 39:43-62.
15. Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens* 2005;26:217-22.
16. Altmann A, Toloşi L, Sander O, et al. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26: 1340-7.
17. Lee C, Zame W, Yoon J, et al. DeepHit: a deep learning approach to survival analysis with competing risks. *Proceedings of the AAAI Conference on Artificial Intelligence* 2018;32.
18. Gradient Boosted Models. at https://scikit-survival.readthedocs.io/en/stable/user_guide/boosting.html.
19. Dorresteijn JAN, Visseren FLJ, Wassink AMJ, et al. Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score. *Heart* 2013;99:866.
20. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997; 30:1145-59.
21. Graham IM, Di Angelantonio E, Visseren F, et al. Systematic coronary risk evaluation (SCORE): JACC focus seminar 4/8. *J Am Coll Cardiol* 2021;77:3046-57.
22. Lloyd-Jones DM, Braun LT, Ndumele CE, et al. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American Heart Association and American College of Cardiology. *Circulation* 2019;139:e1162-77.
23. Bitton A, Gaziano TA. The Framingham Heart Study's impact on global risk assessment. *Prog Cardiovasc Dis* 2010;53:68-78.
24. Selvarajah S, Kaur G, Haniff J, et al. Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population. *Int J Cardiol* 2014;176:211-8.
25. Cho SY, Kim SH, Kang SH, et al. Pre-existing and machine learning-based models for cardiovascular risk prediction. *Sci Rep* 2021; 11:8886.
26. Alaa AM, Bolton T, Di Angelantonio E, et al. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLoS One* 2019;14:e0213653.
27. Wei CC, Lee SH. Predictors of mortality in elderly patients with non-ST elevation acute coronary syndrome - data from Taiwan Acute Coronary Syndrome Full Spectrum Registry. *Acta Cardiol Sin* 2017;33:377-83.
28. Ke J, Chen Y, Wang X, et al. Machine learning-based in-hospital mortality prediction models for patients with acute coronary syndrome. *Am J Emerg Med* 2022;53:127-34.
29. Hadanny A, Shouval R, Wu J, et al. Predicting 30-day mortality after ST elevation myocardial infarction: machine learning-based random forest and its external validation using two independent nationwide datasets. *J Cardiol* 2021;78:439-46.
30. Li X, Liu H, Yang J, et al. Using machine learning models to predict in-hospital mortality for ST-elevation myocardial infarction patients. *Stud Health Technol Inform* 2017;245:476-80.
31. Lee SW, Lee HC, Suh J, et al. Multi-center validation of machine learning model for preoperative prediction of postoperative mortality. *NPJ Digit Med* 2022;5:91.
32. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13:1.
33. Wagner SJ, Matek C, Shetab Boushehri S, et al. Make deep learning algorithms in computational pathology more reproducible and reusable. *Nat Med* 2022;28:1744-6.
34. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
35. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction mod-

els: II. external validation, model updating, and impact assessment. *Heart* 2012;98:691-8.
36. Sylvain A, Alain C. A survey of cross-validation procedures for model selection. *Stat Surv* 2010;4:40-79.
37. Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J King Saud Univ Comput Inf Sci* 2022;34:4514-23.
38. Kulkarni VY, Sinha PK. Pruning of Random Forest classifiers: a survey and future directions. 2012 International Conference on Data Science & Engineering (ICDSE); 18-20 July 2012. p. 64-8.
39. Windeatt T. Accuracy/diversity and ensemble MLP classifier design. *IEEE Trans Neural Netw* 2006;17:1194-211.
40. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019; 110:12-22.
41. Lee C, Light A, Saveliev ES, et al. Developing machine learning algorithms for dynamic estimation of progression during active surveillance for prostate cancer. *NPJ Digit Med* 2022;5:110.
42. Krittanawong C, Virk HUH, Bangalore S, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Sci Rep* 2020;10:16057.
43. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning. *Circ Res* 2017;121:1092-101.
44. Jiang Y, Zhang X, Ma R, et al. Cardiovascular disease prediction by machine learning algorithms based on cytokines in Kazakhs of China. *Clin Epidemiol* 2021;13:417-28.
45. Duarte Lau F, Giugliano RP. Lipoprotein(a) and its significance in cardiovascular disease: a review. *JAMA Cardiol* 2022;7:760-9.
46. Verma KP, Inouye M, Meikle PJ, et al. New cardiovascular risk assessment techniques for primary prevention: JACC review topic of the week. *J Am Coll Cardiol* 2022;80:373-87.
47. Qayyum A, Qadir J, Bilal M, et al. Secure and robust machine learning for healthcare: a survey. *IEEE Rev Biomed Eng* 2021;14: 156-80.
48. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25: 30-6.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Pearson correlation coefficient

	ASCVD	MI	PAOD	Stroke	eGFR	UA	HF	SBP	DM	Smoking	Revasc	LDL-C	Sex	Age	BMI	Statin	TG	Non-HDL-C	Antiplatelet	ARB/ACEi	BB	HDL
ASCVD	1.00																					
MI	0.71	1.00																				
PAOD	0.10	-0.01	1.00																			
Stroke	0.31	-0.05	0.23	1.00																		
eGFR	-0.09	-0.08	-0.06	-0.09	1.00																	
UA	0.12	-0.10	-0.01	-0.03	0.00	1.00																
HF	0.13	0.15	0.02	-0.01	-0.06	0.00	1.00															
SBP	-0.05	-0.05	0.01	0.03	-0.05	0.00	-0.04	1.00														
DM	0.09	0.07	0.02	0.02	-0.08	0.00	0.02	0.07	1.00													
Smoking	0.23	0.22	0.03	0.04	-0.04	0.02	0.06	-0.05	0.03	1.00												
Revasc	0.19	-0.16	0.00	0.00	0.04	-0.03	0.01	0.01	0.04	0.03	1.00											
LDL-C	-0.19	-0.16	-0.02	-0.01	0.06	-0.02	-0.04	0.04	-0.14	-0.06	-0.05	1.00										
Sex	0.29	0.26	0.01	0.03	-0.04	0.02	0.05	-0.06	0.01	0.37	0.07	-0.09	1.00									
Age	0.13	0.05	0.09	0.13	-0.33	0.03	0.11	0.05	0.04	0.00	0.03	-0.11	-0.11	1.00								
BMI	-0.01	-0.01	-0.02	-0.05	0.00	0.02	0.02	0.09	0.13	0.02	0.02	-0.02	0.03	-0.20	1.00							
Statin	0.21	0.28	-0.04	-0.07	-0.03	0.02	0.03	-0.05	0.07	0.06	0.05	-0.11	0.02	0.00	0.04	1.00						
TG	-0.02	0.00	-0.02	-0.02	-0.04	0.00	0.01	0.05	0.12	-0.01	0.00	-0.02	0.04	-0.17	0.15	-0.02	1.00					
Non-HDL-C	-0.19	-0.16	-0.03	-0.01	0.02	-0.02	-0.04	0.06	-0.08	-0.06	-0.05	0.85	-0.07	-0.17	0.05	-0.11	0.36	1.00				
Antiplatelet	0.49	0.42	0.01	0.10	-0.08	0.05	0.07	-0.05	0.09	0.13	0.10	-0.11	0.19	0.13	0.00	0.26	0.01	-0.10	1.00			
ARB/ACEi	0.03	0.04	0.00	0.00	-0.06	-0.03	0.05	0.09	0.09	0.01	-0.02	-0.04	0.02	0.04	0.11	0.06	0.05	-0.01	0.08	1.00		
BB	0.07	0.12	-0.06	-0.10	0.02	0.01	0.05	0.01	0.04	0.01	0.07	-0.07	0.04	-0.06	0.09	0.09	0.06	-0.03	0.06	0.02	1.00	
HDL	-0.17	-0.15	0.01	-0.01	0.02	-0.01	-0.03	-0.01	-0.15	-0.12	-0.06	0.09	-0.29	0.11	-0.17	0.02	-0.30	-0.08	-0.09	-0.06	-0.12	1.00

ARB/ACEi, angiotensin receptor blockers/angiotensin-converting enzyme inhibitors; ASCVD, atherosclerotic cardiovascular disease; BMI, body mass index; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; LDL-C, Low-density lipoprotein cholesterol; PAOD, peripheral arterial occlusive disease; SBP, systolic blood pressure; TG, triglyceride.