



OPEN ACCESS

EDITED BY

Yanan Wu,
China Medical University, China

REVIEWED BY

Jiayin Zhang,
Shanghai General Hospital, China
Liqing Peng,
Sichuan University, China

*CORRESPONDENCE

Ting Liu
✉ cmuliuting@sina.cn
Hong Shang
✉ hongshang100@hotmail.com

†These authors have contributed equally to this work and share first authorship

†These authors have contributed equally to this work and share senior authorship

RECEIVED 08 July 2025

ACCEPTED 25 August 2025

PUBLISHED 16 September 2025

CITATION

Liu H, Ding H, Zheng Y, Li Y, Yu Y, Geng Z, Zhou J, Huo H, Li H, Peng X, Tian Z, Li X, Shang H and Liu T (2025) Prediction of obstructive coronary artery disease in people living with HIV: value of machine learning incorporating HIV-specific factors. *Front. Med.* 12:1661990. doi: 10.3389/fmed.2025.1661990

COPYRIGHT

© 2025 Liu, Ding, Zheng, Li, Yu, Geng, Zhou, Huo, Li, Peng, Tian, Li, Shang and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prediction of obstructive coronary artery disease in people living with HIV: value of machine learning incorporating HIV-specific factors

Hui Liu^{1,2†}, Haibo Ding^{3,4,5†}, Yue Zheng¹, Yue Li¹, Yang Yu⁶, Zhaodi Geng¹, Jie Zhou¹, Huaibi Huo¹, Han Li⁷, Xin Peng^{1,8}, Zhaoxin Tian¹, Xiaolin Li^{3,4,5}, Hong Shang^{3,4,5*†} and Ting Liu^{1*†}

¹Department of Radiology, The First Hospital of China Medical University, Shenyang, China,

²Department of CT, Shaanxi Provincial People's Hospital, Xi'an, China, ³State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, NHC Key Laboratory of AIDS Prevention and Treatment, National Clinical Research Center for Laboratory Medicine, The First Hospital of China Medical University, China Medical University, Shenyang, China, ⁴Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, China, ⁵Key Laboratory of AIDS Immunology of Liaoning Province, Shenyang, China, ⁶Siemens Healthineers China, Shanghai, China, ⁷Department of Radiology, Jing'an District Centre Hospital of Shanghai, Fudan University, Shanghai, China,

⁸Department of Radiology, The Third People's Hospital of Chengdu, Chengdu, China

Objectives: To explore the value of machine learning (ML) model in conjunction with HIV-specific risk factors to predict obstructive coronary artery disease (CAD) ($\geq 50\%$ stenosis) on coronary CT angiography (CTA) in the asymptomatic people living with HIV (PLWH).

Methods: In this cross-sectional study, we prospectively analyzed 304 PLWH without chest pain (age 48 ± 11 years, 91% males). The dataset was randomly divided into training and held-out test sets in an 8:2 ratio. The ML model established by random forest was compared with traditional models, including CAD consortium clinical score, CONFIRM score, and Genders clinical model, as well as logistic regression model. The coronary artery calcium score (CACS) was added to the above five models to establish new models. Predictive performance of the models was evaluated according to Delong test.

Results: Obstructive CAD occurred in 64 of 304 PLWH (21%). The ML model (AUC of 0.946) had the highest discrimination for obstructive CAD compared with above models (AUC of 0.734, 0.736, 0.737, and 0.782, respectively; $p < 0.05$ for all comparisons). ML model showed the best calibration and clinical decision-making capability. Moreover, the ML model showed the best predictive performance compared with models after adding the CACS (AUC of 0.772, 0.740, 0.742, 0.750, and 0.798, respectively; $p < 0.05$ for all comparisons).

Conclusion: The ML model incorporating cardiovascular risk factors and HIV-specific factors can more accurately estimate the pretest likelihood of obstructive CAD in PLWH than traditional models. ML improves risk stratification in HIV populations and may help guide management.

KEYWORDS

HIV, coronary artery disease, pre-test probability, coronary CTA, machine learning

Introduction

The AIDS epidemic has been a global public health issue for more than 40 years and has resulted in approximately 84 million infections and 40 million deaths (1). Antiretroviral therapy (ART) can slow the progression of the disease and prolong the life of an infected person by suppression of viral replication and immune restoration (2, 3). Persistent immune activation despite ART, HIV-specific risk factors, and the aging PLWH population have been associated with a higher risk of cardiovascular disease (CAD), compared with the general population (4–6).

Coronary CT angiography (CTA) is widely accepted and has become the main imaging test for the diagnosis of CAD (7). However, in daily clinical practice, a significant number of individuals undergoing coronary CTA have minimal or no CAD, which results in additional radiation exposure and suboptimal cost-effectiveness (8). The European guideline has therefore recommended the performance of pre-test probability assessments to guide clinical decisions on whether diagnostic testing should be deferred or conducted and whether the initial test should be non-invasive or invasive (9).

Existing conventional prediction models of obstructive CAD, including the CAD consortium clinical score, CONFIRM registry score (CRS), and genders clinical model (GCM) are only applicable to non-HIV populations with chest pain (9–11). However, these general population-derived models are not directly applicable to the asymptomatic high-risk PLWH population and HIV-specific risk factors contributing to CAD.

Artificial intelligence and machine learning (ML) have shown promise to improve risk assessment in various clinical scenarios, with iterative algorithms having an advantage over linear algorithms for the internal relations among confounding factors. Applications of ML have not only been limited to medical image reading, but have also been widely used in medical data analysis (12). Furthermore, the integrated use of clinical and imaging data to predict clinical outcomes shows great promise (13).

Accordingly, we sought to explore the value of a machine learning model combining cardiovascular risk factors, hematologic indicators, and HIV-specific factors to predict the risk of obstructive CAD in PLWH more accurately.

Materials and methods

Study population

In this cross-sectional study, we enrolled 346 PLWH who underwent 64-detector row coronary CTA evaluation between December 2019 and June 2023. The inclusion criteria for PLWH included (1) diagnosis of HIV infection, (2) age between 20 and 80 years, (3) undergoing ART. The exclusion criteria included: (1) previous history or symptoms CAD, (2) contraindications of CCTA

examination, (3) poor quality of coronary images, and (4) not have complete clinical information (Figure 1). The final dataset contained no missing values. Eventually, 304 patients were enrolled for final analysis. Fifty of these patients have been used for the earlier MRI study, but the current CT study is analyzing more patients with a completely different goal and result (14). This study was approved by the local Ethics Review Board and conducted in accordance with the Declaration of Helsinki.

Data collection

Prior to the coronary CTA examination, demography (age, sex), clinical information (cardiovascular risk factors, antihypertensive therapy, HIV acquisition mode, hepatitis C, intravenous drug use, the time since HIV diagnosis), hematologic and immunological factors (estimated glomerular filtration rate, HIV viral load, current CD4⁺ T-cell count, CD8⁺ T-cell count, and nadir CD4⁺ T-cell count, CD4⁺/CD8⁺), and treatment-related factors (duration on ART, ART regimen) was prospectively collected for all PLWH participants.

Coronary CTA scanning protocol and image analysis

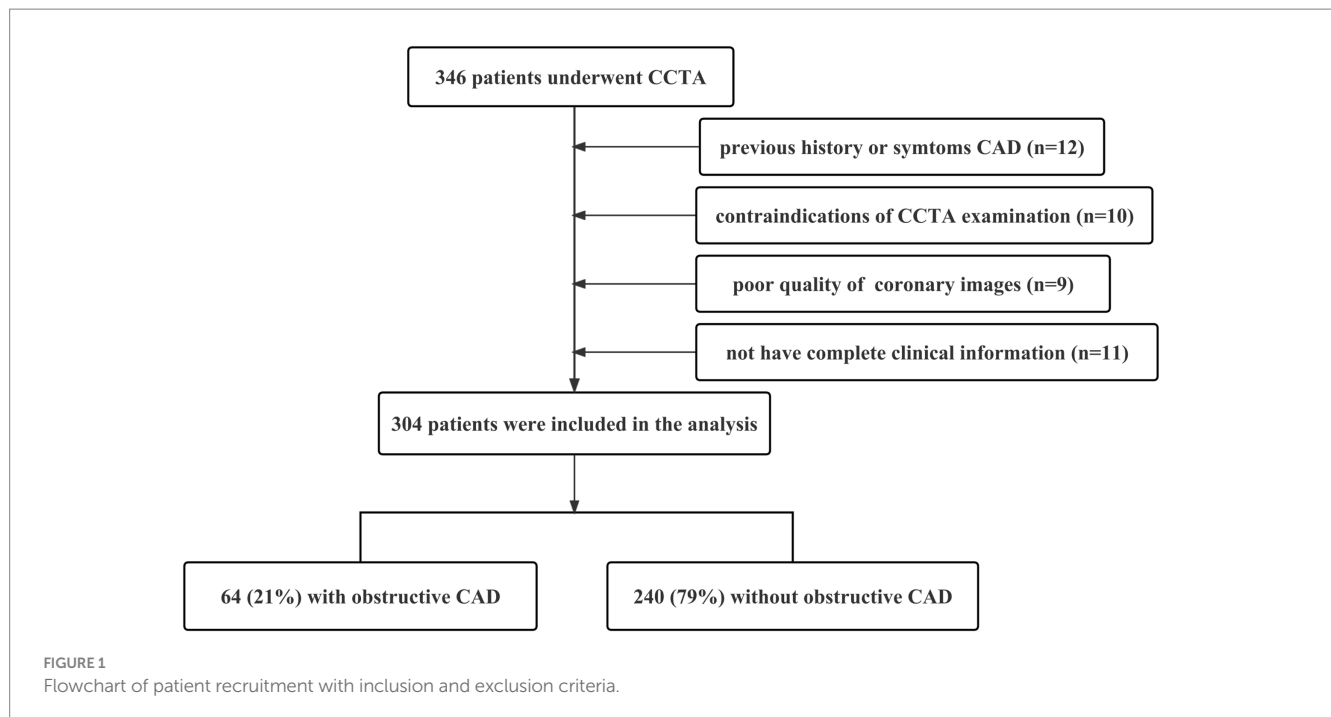
Coronary CTA was performed using a third-generation dual-source CT (SOMATOM Force, Siemens Healthcare, Forchheim, Germany) and 45 to 100 mL of intravenous contrast (Iopamidol Injection, 370 mg/mL, Bracco Si, Shanghai, China) at a flow rate of 4.0–5.0 mL/s. Retrospective electrocardiographic gating was used in all of the participants. The main scan parameters were as follows: tube voltage = 100 kV, tube current = 300–400 mas, slice thickness = 0.75 mm, detector collimation = 2 mm × 64 mm × 0.6 mm, and matrix = 512 × 512. A non-contrast CT examination was performed for coronary calcium scoring.

All the scans were sequentially analyzed and reviewed on a workstation (Syngo.via version VB20, Siemens Healthineer, Forchheim, Germany) by two radiologists (with three and fifteen years of experience, respectively) using axial images and multiplanar reconstructions. The coronary segment was analyzed by using the 16-segment American Heart Association definitions from 1975 (15). The outcome was the presence of obstructive CAD demonstrated by coronary CTA, which has been shown in previous studies to have a high diagnostic accuracy for obstructive CAD (16, 17). Obstructive CAD was defined as at least 1 coronary segment with a lesion of ≥50% luminal stenosis in diameter (18). Coronary artery calcium score (CACS) measurement was performed using the Agatston method (19). Observations between the two readers were analyzed for interobserver agreement.

Model construction

We developed a total of 11 models to predict obstructive CAD. There are three traditional conventional pre-test probability models of coronary CTA. All our patients were asymptomatic, so clinical symptoms could not be included as a feature in the

Abbreviations: CAD, Coronary artery disease; PLWH, People living with HIV; ML, Machine learning; RF, Random forest; LR, Logistic regression; AUC, Area under the curve; ART, Antiretroviral therapy; CRS, CONFIRM Registry Score; GCM, Genders clinical model.



traditional model. In addition to gender, age, hypertension, diabetes, and smoking status, CAD consortium clinical score requires hyperlipidemia and BMI, CRS requires cardiovascular family history, and GCM requires hyperlipidemia. The logistic regression (LR) model was built using all available features, including traditional cardiovascular and HIV-specific variables. Moreover, we added CACS to above four models to create four additional models: CAD + CACS, CRS + CACS, GCM + CACS, and LR + CACS. A CACS model using only the calcium score was included for comparison.

Furthermore, three ML models were developed based on the random forest algorithm. Model 1 incorporated all available clinical, laboratory, and HIV-specific variables. After adding CACS, Model 2 was reconstructed with renewed feature selection and hyperparameter tuning. To evaluate the incremental value of CACS, Model 3 retained the features and hyperparameters selected in Model 1, with CACS added as an additional variable.

Consistent with the 2019 ESC Guidelines, our machine learning model incorporates multiple cardiovascular risk factors to estimate the likelihood of obstructive CAD (20). To facilitate clinical risk interpretation, we adopted the risk stratification thresholds proposed in the 2013 ESC Guidelines based on pretest probability: low risk (<15%), intermediate risk (15–85%), and high risk (>85%) (21). Patients with a low pretest probability are less likely to benefit from additional testing, while those with an intermediate probability are most likely to benefit from initial noninvasive testing.

Machine learning

Random forest algorithm was performed to model our data. The original dataset was randomly divided into training and

held-out test sets in an 8:2 ratio to maintain the ratio of obstructive to non-obstructive CAD in both subsets. In addition, model hyperparameters are optimized by five-fold cross-validation, which repeats five iterations by randomly splitting the training set into five equal-sized subsets, four were used for inner training and one was used for inner test. The employ of cross-validation in the training set can be beneficial as its primary use is to empirically determine the optimal model hyperparameters without resorting to the test set (22). Subsequently, the optimal hyperparameters are applied to evaluate the model in the test set (maximum tree depth: 4; number of trees: 5; random seed: 2175). Finally, the area under the curve (AUC) and the associated 95% confidence interval were used to measure the classification performance of the ML model. Feature selection is performed using the Boruta algorithm. The feature selection and modeling process were performed in the training set, and the independent test set was only used to evaluate the performance of the model and was not involved in the training process.

Statistical analysis

Continuous variables were expressed as median with 25th–75th percentile interquartile range (IQR), depending on the normality of distribution assessed by the Kolmogorov–Smirnov test. Categorical variables were presented as frequencies and percentages. Comparisons between groups were carried out using an independent t-test or the Mann–Whitney U test for normally and non-normally distributed continuous variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

The discriminative power of Model 1 was compared with the 4 traditional models and LR model, respectively, using

receiver-operating characteristic analysis and pairwise comparisons according to Delong et al. (23). $p < 0.05$ indicated statistical significance between the performance of the models. Calibration curves were drawn to explore the agreement between the observed outcome frequencies and predicted probabilities of the models in the testing group. Decision curve analysis was used to assess the clinical utility of these models by quantifying the net benefits at different risk threshold probabilities in testing set.

All statistical analyses were carried out using IBM SPSS software (V26.0; IBM Corporation, Armonk, New York, USA) and R software (version 4.3.1).

Results

A total of 304 patients met the inclusion criteria and were included in the analysis. The mean age of the patients was 48 ± 10 years, and 91% were male. The occurrence of obstructive CAD was 21% (64/304) within the studied cohort (Figure 2). The clinical characteristics are listed in Table 1. The presence of obstructive CAD was significantly associated with age, hypertension, use of antihypertensive drugs, total cholesterol, low-density lipoprotein, hyperlipidemia, and estimated glomerular filtration rate. Among HIV-related indicators, nadir CD4⁺ T-cell count and duration of antiretroviral drug therapy were significantly associated with obstructive CAD.

The importance of features

Figure 3 presents the feature importance rankings generated by the Boruta algorithm. As shown, age, total cholesterol, and current

CD4⁺ T-cell count were the most predictive features in the Model 1. This was followed by years of ART use, triglycerides, nadir CD4⁺ T-cell count, and diastolic blood pressure. Interestingly, the importance of traditional cardiovascular risk factors increased with the addition of the CACS into the model, the most predictive features (after the CACS itself) were triglycerides and age followed by estimated glomerular filtration rate and CD4⁺/CD8⁺. The risk factors for obtaining models after screening are shown in Table 2.

Discriminative power

The machine learning model (Model 1, AUC of 0.946), which incorporated both traditional cardiovascular risk factors and HIV-specific metrics, showed the best performance in predicting obstructive CAD compared with CAD Consortium clinical score, CRS, and GCM (AUC of 0.734, 0.736, and 0.737, respectively; $p < 0.05$ for all comparisons). In order to control for potential confounding due to algorithm differences, we additionally constructed a logistic regression (LR) model using the full feature set (LR model, AUC of 0.782; $p < 0.05$), confirming the superior predictive capability of the random forest-based Model 1 ($p < 0.05$). Furthermore, the LR model did not differ significantly from traditional models. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for the prediction of obstructive CAD were 0.846, 0.958, 0.846, 0.958, and 0.934 for Model 1, respectively. Interestingly, Model 2, developed by adding CACS and reapplying feature selection, yielded a slightly lower AUC of 0.926 but still outperformed the CAD + CACS model, CRS + CACS model, GCM + CACS model, and LR + CACS model (AUC of 0.740, 0.742, 0.750, and 0.798; $p < 0.05$ for all comparisons). After adding CACS to the Model 1 feature set without re-optimization, Model 3 demonstrated a reduced AUC of 0.890, with no statistically significant

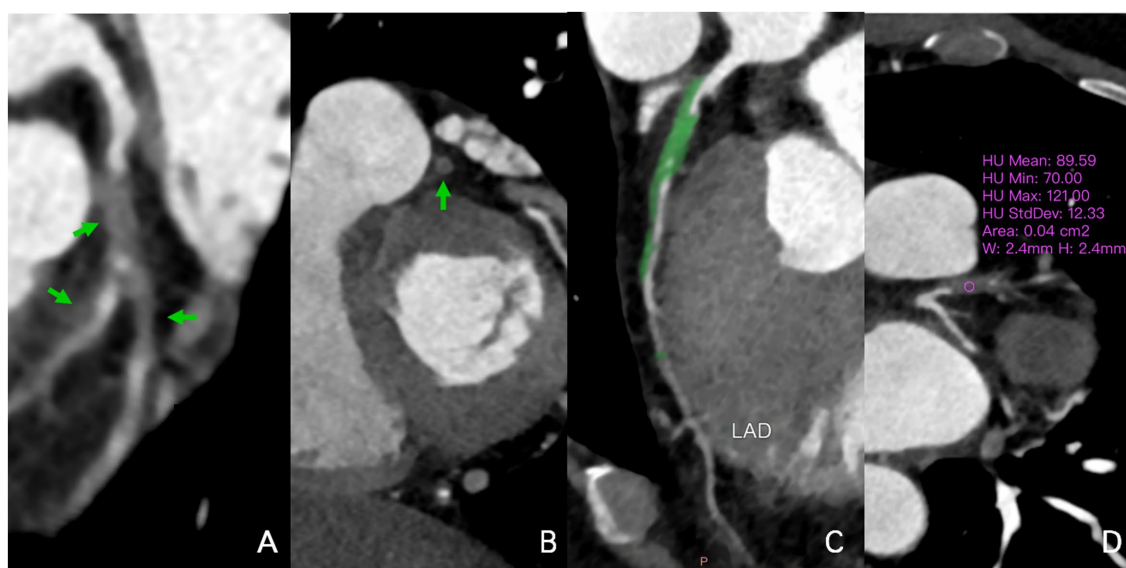


FIGURE 2

A 58-year-old man infected with HIV. In the multiplanar reconstruction of coronary CTA (A–D), non-calcified plaques (70–121HU) were seen in the proximal segment of the LAD and the beginning of the first diagonal branch, with severe lumen stenosis (90–99%). Noncalcified plaque also was seen in the middle segment of the LAD with mild lumen stenosis (40–50%). HIV, human immunodeficiency virus; CTA, CT angiography; HU, Hounsfield units; LAD, left anterior descending artery.

TABLE 1 Patient characteristics.

Characteristic	All participants (<i>n</i> = 304)	With obstructive CAD (<i>n</i> = 64)	Without obstructive CAD (<i>n</i> = 240)	<i>p</i> -value
Male sex, <i>n</i> (%)	277 (91.1)	60 (93.8)	217 (90.4)	0.405
Age (years), median (IQR)	49.0 (41.0–56.0)	55.5 (49.0–61.0)	47.0 (40.0–53.0)	<0.001
Body mass index (kg/m ²), median (IQR)	23.0 (21.5–25.3)	23.2 (21.7–25.5)	22.9 (21.4–25.2)	0.552
Systolic pressure (mmHg), median (IQR)	125.0 (115.0–135.0)	135.0 (121.0–144.8)	126.0 (120.0–138.0)	0.001
Diastolic pressure (mmHg), median (IQR)	80.0 (78.0–90.0)	88.0 (80.0–96.0)	81.5 (78.0–90.0)	0.001
Hypertension, <i>n</i> (%)	39 (12.8)	14 (21.9)	25 (10.4)	0.015
Antihypertensive drugs, <i>n</i> (%)	26 (8.6)	10 (15.6)	16 (6.7)	0.023
Diabetes mellitus, <i>n</i> (%)	20 (6.6)	7 (10.9)	13 (5.4)	0.113
Total cholesterol (mmol/L), median (IQR)	4.7 (4.1–5.3)	4.9 (4.5–5.7)	4.6 (4.0–5.2)	0.004
HDL cholesterol (mmol/L), median (IQR)	1.1 (1.0–1.3)	1.1 (0.9–1.3)	1.2 (1.0–1.4)	0.971
LDL cholesterol (mmol/L), median (IQR)	2.8 (2.3–3.5)	2.9 (2.2–3.8)	2.8 (2.3–3.4)	0.103
Triglyceride (mmol/L), median (IQR)	1.8 (1.1–2.7)	1.9 (1.3–3.4)	1.8 (1.1–2.6)	0.112
Hyperlipidemia, <i>n</i> (%)	193 (63.5)	48 (75.0)	145 (60.4)	0.031
Current smoking, <i>n</i> (%)	97 (31.9)	23 (35.9)	74 (30.8)	0.436
Current alcohol, <i>n</i> (%)	114 (37.5)	25 (39.1)	89 (37.1)	0.771
Family history of CAD, <i>n</i> (%)	70 (23.0)	18 (28.1)	52 (21.7)	0.276
eGFR, median (IQR)	106.7 (98.2–115.1)	99.6 (88.3–107.6)	108.0 (99.8–114.3)	<0.001
Hepatitis C seropositivity, <i>n</i> (%)	4 (1.3)	1 (1.6)	3 (1.3)	1.000
CACS, median (IQR)	0 (0–3.4)	12.4 (0–82.5)	0	<0.001
CD4 ⁺ count (cells/mm ³), median (IQR)	591.5 (455.5–768.0)	580.5 (410.8–713.0)	595.0 (458.0–770.5)	0.199
CD4 ⁺ nadir (cells/mL ³), median (IQR)	283.2 (177.0–377.4)	220.2 (92.5–363.3)	295.5 (201.0–378.9)	0.032
CD4 ⁺ /CD8 ⁺ , median (IQR)	0.8 (0.6–1.2)	0.7 (0.5–1.0)	0.8 (0.6–1.2)	0.618
Active illicit drug use, <i>n</i> (%)	5 (1.6)	3 (4.7)	2 (0.8)	0.065
On ART, <i>n</i> (%)				0.242
NNRTI	194 (63.8)	39 (60.9)	155 (64.6)	
IN	87 (28.6)	17 (26.6)	70 (29.2)	
PI	23 (7.6)	8 (12.5)	15 (6.3)	
Years of HIV-diagnosed, median (IQR)	7.0 (4.0–11.0)	8.0 (4.3–10.8)	7.0 (4.0–11.0)	0.577
Years of HIV-treated, median (IQR)	8.0 (5.0–11.0)	8.0 (5.0–11.0)	7.0 (5.0–9.0)	0.208
Viral load ≥50 copies/mL, <i>n</i> (%)	34 (11.2)	11 (17.2)	23 (9.6)	0.086

IQR, interquartile range; HDL cholesterol, high density lipoprotein cholesterol; LDL cholesterol, low density lipoprotein cholesterol; CAD, cardiovascular disease; eGFR: estimated glomerular filtration rate; CACS, coronary artery calcium score; ART, antiretroviral therapy; NNRTI, non-nucleoside reverse-transcriptase inhibitor; IN, integrase inhibitors; PI, protease inhibitors.

difference compared to Model 1 or Model 2 ($p > 0.05$). Meanwhile, the performance of Model 1 and Model 2 significantly outperformed the CACS models (AUC of 0.776; $p < 0.05$), suggesting that clinical and HIV-specific features contributed more substantially to predictive performance than CACS alone. The discriminative performance of all models is shown in [Tables 3, 4](#) and [Figure 4](#).

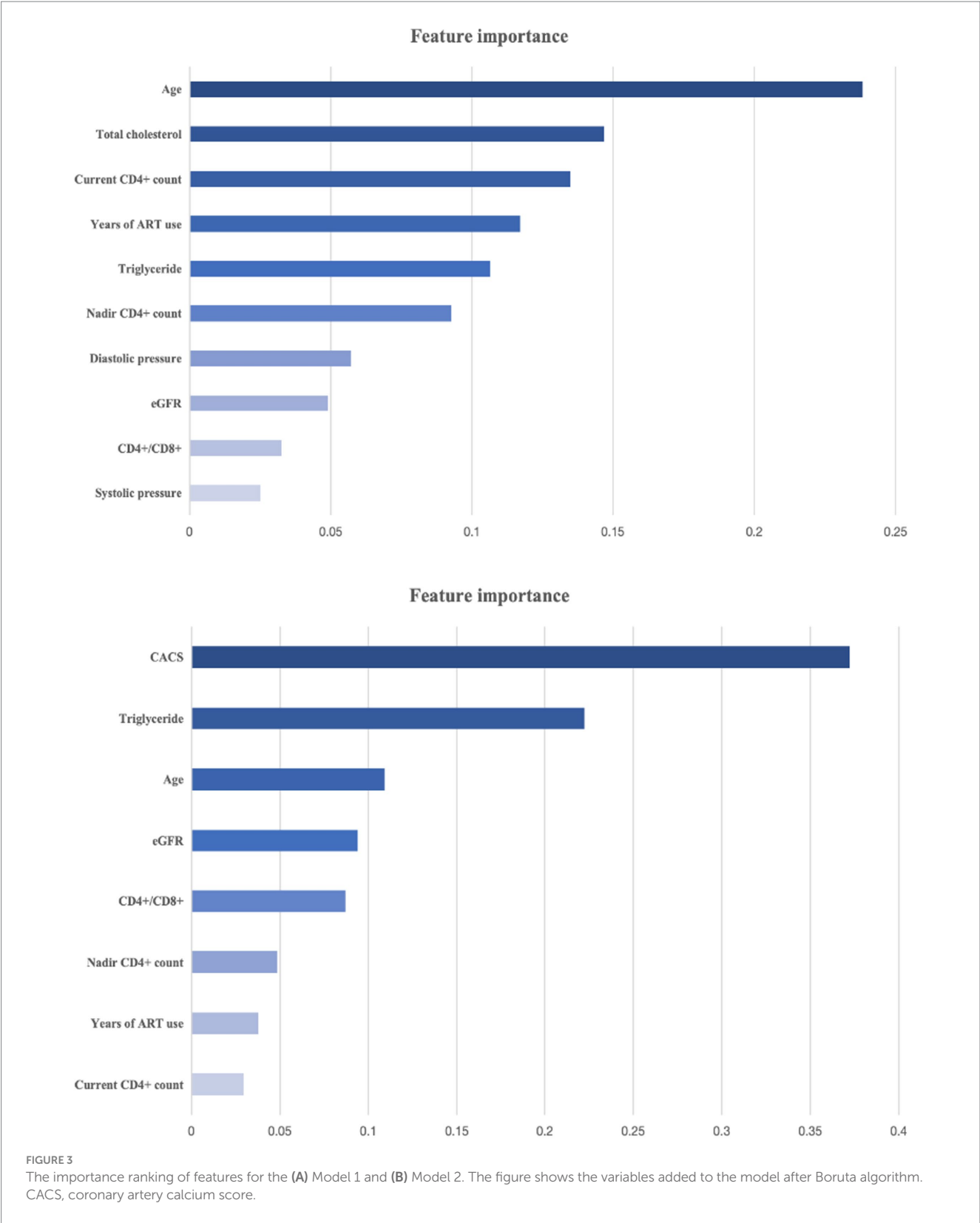
Calibration and clinical decision

In both groups, the calibration curves of Model 1 and Model 2 showed the best calibration of the predicted probability with the true probability (Brier scores of 0.076 and 0.102, respectively) ([Figure 5](#)). The clinical usefulness is shown as decision curve in [Figure 6](#). The ML Models demonstrated superior net benefit in

decision curve analysis when the threshold probability ranged from 0 to 70%.

Discussion

This cross-sectional study explores the prediction of the probability of obstructive CAD in a cohort of PLWH by using machine learning in conjunction with clinical risk factors and HIV-specific risk markers. We found that HIV-associated immunologic markers are strongly associated with predicting the prevalence of obstructive CAD in populations with well-controlled HIV disease. This study extends our understanding of obstructive CAD in PLWH, demonstrating the improved performance of machine learning models constructed with the random forest algorithm.



We demonstrated that machine learning models were more applicable to HIV-infected individuals than traditional models with only traditional risk factors. Existing conventional prediction models are only used in the symptomatic general population, with the prevalence of obstructive CAD ranging from 18–24% in previous studies (24–26). However, in this study, stenosis greater than 50% in asymptomatic HIV patients accounted for 21%, indicating some of the PLWH had already suffered from obstructive CAD during the subclinical period without

timely examination and further treatment, which may lead to major adverse cardiac events. Therefore, early diagnosis and treatment targeting high-risk populations hold great clinical relevance. However, the use of the LR algorithm did not improve the performance of the model even when HIV-specific factors were added. This suggests that there is a non-linear relationship among HIV-specific factors, and that

TABLE 2 The risk factors for obtaining models.

Models	Features
CAD	age, gender, diabetes, hypertension, hyperlipidemia, smoking, BMI
CRS	age, gender, diabetes, hypertension, smoking, family history of CVD
GCM	age, gender, diabetes, hypertension, smoking, hyperlipidemia
LR Model	age, TG, nadir CD4 ⁺ T-cell count, diastolic pressure
CACS Model	CACS
CAD + CACS	CACS, age, gender, diabetes, hypertension, hyperlipidemia, smoking, BMI
CRS + CACS	CACS, age, gender, diabetes, hypertension, smoking, family history of CVD
GCM + CACS	CACS, age, gender, diabetes, hypertension, smoking, hyperlipidemia
LR + CACS	CACS, age, TG, nadir CD4 ⁺ T-cell count, diastolic pressure
Model 1	age, TC, current CD4 ⁺ T-cell count, years of ART use, TG, nadir CD4 ⁺ T-cell count, Diastolic blood, eGFR, CD4 ⁺ /CD8 ⁺ , systolic pressure
Model 2	CACS, TG, age, CD4 ⁺ /CD8 ⁺ , nadir CD4 ⁺ T-cell count, years of ART use, current CD4 ⁺ T-cell count
Model 3	CACS, age, TC, current CD4 ⁺ T-cell count, years of ART use, TG, nadir CD4 ⁺ T-cell count, Diastolic blood, eGFR, CD4 ⁺ /CD8 ⁺ , systolic pressure

CAD: CAD Consortium clinic score; CRS, CONFIRM score; GCM, Genders clinical model; LR, logistics regression; BMI, body mass index; CVD, cardiovascular disease; TG, triglyceride; TC, total cholesterol; ART, antiretroviral therapy; eGFR: estimated glomerular filtration rate; CACS, coronary artery calcium score.

logistic algorithms alone are unable to make better predictions of outcomes resulting from complex and multiple factors (27, 28). Random forests not only explore HIV-specific markers that have an impact on CAD occurrence through an iterative algorithm, but also have a higher predictive ability for obstructive CAD than traditional or LR models (AUC value of 0.946 and accuracy of 0.934).

The mechanisms underlying the increased prevalence of CAD in PLWH are not fully understood, but traditional risk factors are thought to explain only part of the increased risk (29, 30). Among HIV-infected individuals, the presence of obstructive CAD was associated with lower nadir CD4⁺ T-cell count and longer treatment with ART, similar to previous findings (31). They are both markers of longer duration of HIV infection with potential adverse metabolic effects. In addition, studies have shown that while antiretroviral drugs inhibited viral replication and led to increased longevity in HIV populations, they also increased cardiovascular risk (32, 33). In addition, the current CD4⁺ T-cell counts are used to monitor HIV infection status and the efficacy of ART (5), emerged as the third most predictive feature after age and total cholesterol. Furthermore, the CD4⁺ count < 500 cells/mm³ has been shown to be an independent risk factor for CAD, and its attributable risk is equivalent to traditional CAD risk factors (34). In some previous studies, the CD4⁺: CD8⁺ ratio, a marker of immune senescence, has been predictive of cardiovascular events, which is consistent with our study (35, 36).

The model containing only CACS (CACS model) had limited performance in predicting the prevalence of obstructive CAD in HIV populations. Compared with traditional models after adding the CACS, the Model 1 still showed the highest discriminative performance, which suggested that the Model 1 was able to reclassify CAD risk more accurately the PLWH population at low to moderate risk for obstructive CAD using only clinical information, avoiding the use of additional CACS. To further explore the incremental value of CACS, we constructed Model 3 by retaining the selected features and hyperparameters of Model 1 and adding CACS. The results showed that the model performance did not improve, with the AUC declining slightly and no statistically significant difference observed. These may be attributed to the relatively low burden of calcified plaques in asymptomatic PLWH

TABLE 3 The Discriminative Power of all Models for Obstructive CAD.

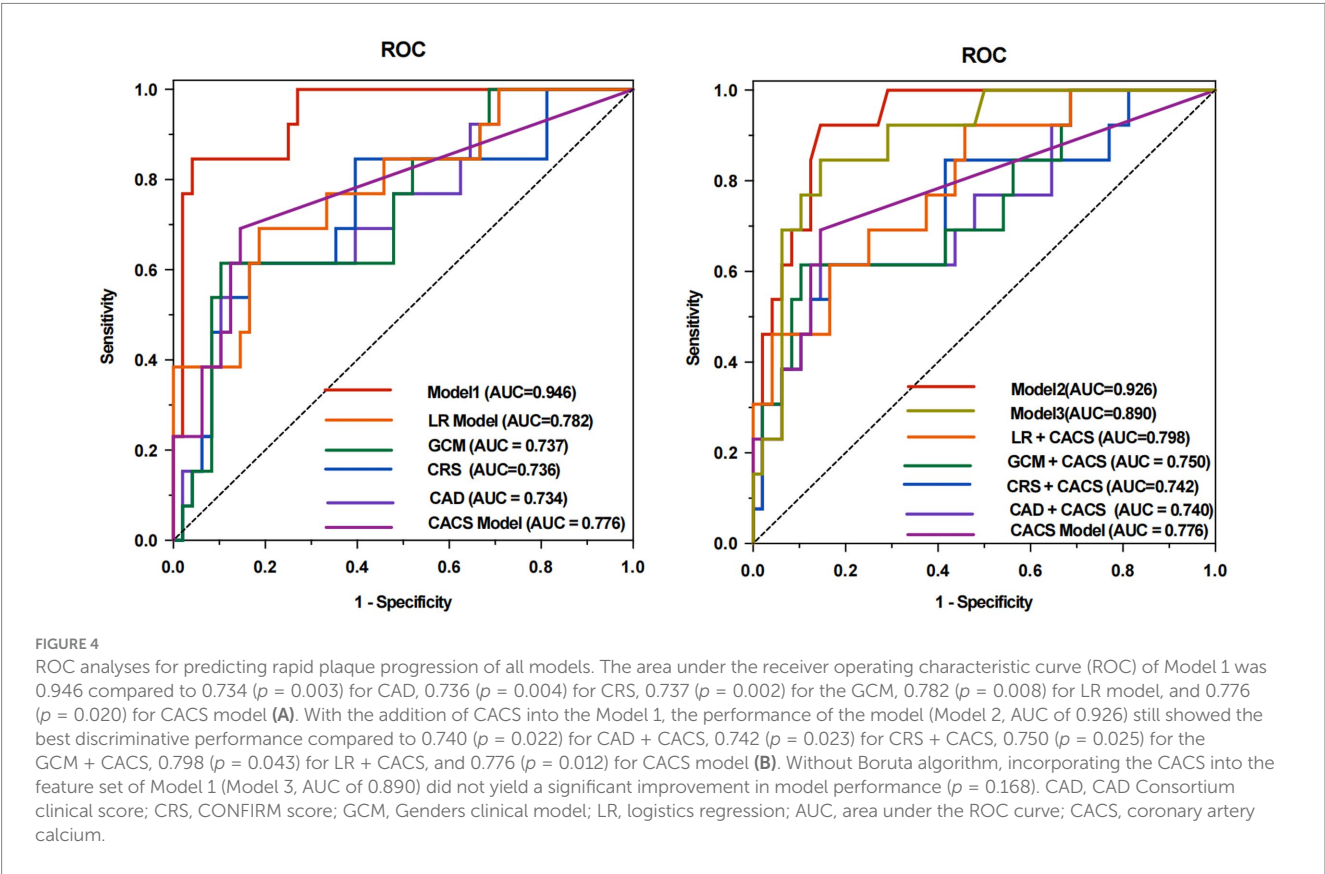
Models	Statistical approach	AUC (95% CI)	p-value (compare with Model1)	p-value (compare with Model2)
CAD	Logistic regression	0.734 (0.605–0.839)	0.003	0.018
CRS	Logistic regression	0.736 (0.607–0.840)	0.004	0.023
GCM	Logistic regression	0.737 (0.609–0.842)	0.002	0.016
LR Model	Logistic regression	0.782 (0.658–0.878)	0.008	0.045
CACS Model	Logistic regression	0.776 (0.651–0.873)	0.020	0.012
CAD + CACS	Logistic regression	0.740 (0.612–0.844)	0.005	0.022
CRS + CACS	Logistic regression	0.742 (0.614–0.846)	0.006	0.023
GCM + CACS	Logistic regression	0.750 (0.623–0.852)	0.006	0.025
LR + CACS	Logistic regression	0.798 (0.676–0.890)	0.005	0.043
Model 1	Random forest	0.946 (0.856–0.987)	\	0.571
Model 2	Random forest	0.926 (0.830–0.977)	0.571	\
Model 3	Random forest	0.890 (0.799–0.969)	0.168	0.285

CAD, CAD Consortium clinic score; CRS, CONFIRM score; GCM, Genders clinical model; LR, logistics regression; AUC, area under the ROC curve; CI, confidence interval; CACS, coronary artery calcium score.

TABLE 4 The performance of all models in the training and testing groups.

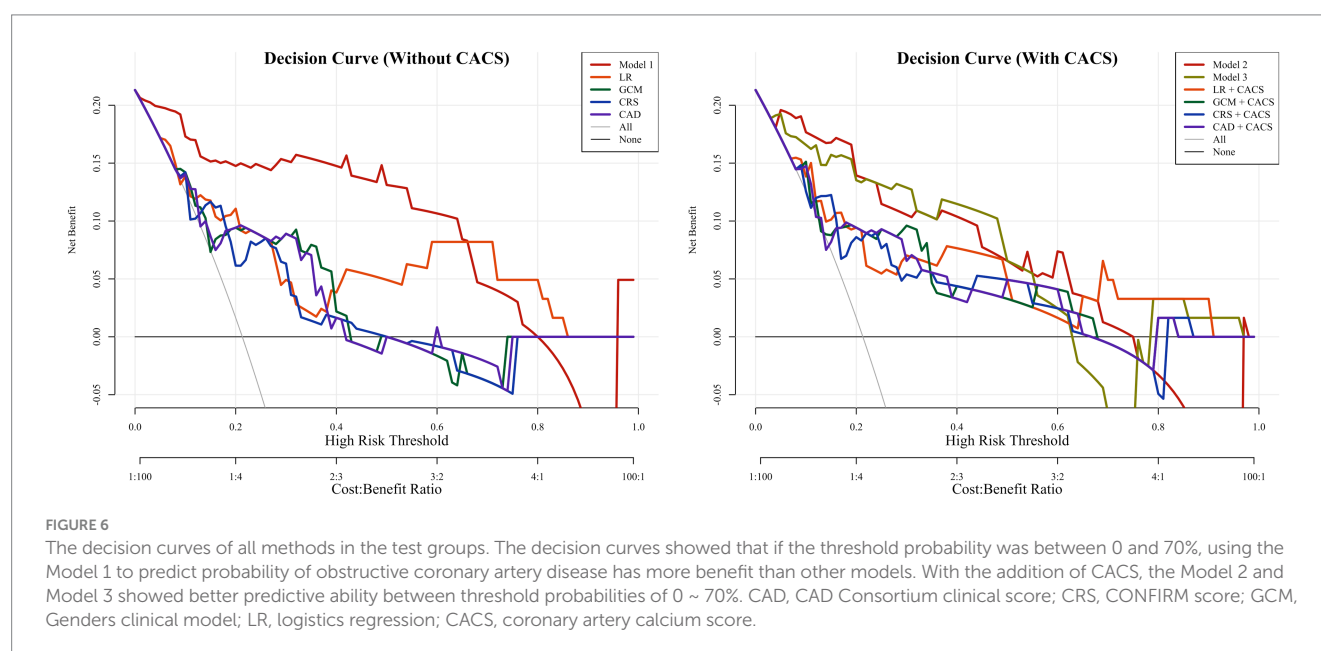
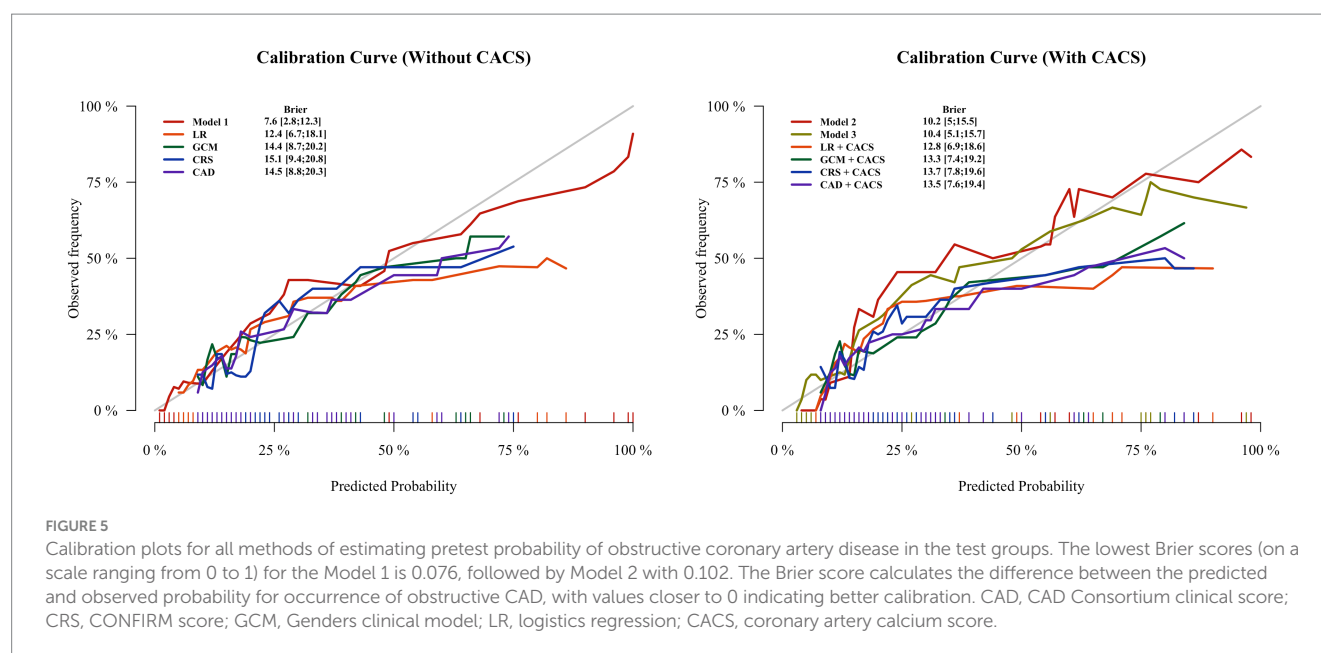
Model	Training group					Testing group				
	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV
Model 1	0.868	0.882	0.865	0.634	0.965	0.934	0.846	0.958	0.846	0.958
Model 2	0.926	0.804	0.958	0.837	0.948	0.869	0.923	0.854	0.632	0.976
Model 3	0.926	0.706	0.984	0.923	0.927	0.853	0.539	0.938	0.700	0.882
LR Model	0.757	0.686	0.776	0.449	0.903	0.787	0.692	0.813	0.500	0.907
GCM	0.798	0.490	0.880	0.521	0.867	0.820	0.615	0.875	0.571	0.894
CRS	0.757	0.698	0.797	0.443	0.884	0.771	0.692	0.792	0.474	0.905
CAD	0.811	0.412	0.917	0.568	0.854	0.836	0.615	0.896	0.615	0.896
CACS	0.778	0.667	0.807	0.479	0.901	0.820	0.692	0.854	0.562	0.911
LR + CACS	0.807	0.725	0.828	0.529	0.919	0.623	0.923	0.542	0.353	0.963
GCM + CACS	0.720	0.882	0.677	0.421	0.956	0.836	0.615	0.896	0.615	0.896
CRS + CACS	0.794	0.706	0.818	0.507	0.913	0.787	0.615	0.833	0.500	0.889
CAD + CACS	0.753	0.784	0.745	0.449	0.929	0.803	0.615	0.854	0.533	0.891

ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value. CAD, CAD Consortium clinic score; CRS, CONFIRM score; GCM, Genders clinical model; LR, logistics regression; CACS, coronary artery calcium score.



(37), as well as potential collinearity between CACS and other variables such as age and lipid profile. These findings suggest that although CACS may be associated with CAD risk, it does not necessarily provide additional predictive value beyond a well-optimized model incorporating both traditional and HIV-specific clinical features. However, this finding still needs to be supported by prospective randomized trials with a larger sample size that could focus on assessing the importance of the coronary calcium score in the ML model.

CT contrast agents are more likely to produce drug allergic reactions in PLWH participants than the general population due to the continuous stimulation of immune cells in the HIV population (38). However, previous pretest probability models, including the CAD Consortium scores, GCM, and CRS, have had limited predictive power for HIV populations. The Model 1 with significantly higher AUC values performs well in terms of accuracy, calibration, and net-benefit (risk threshold from 0 to



80%). As a pre-test assessment (the clinical likelihood that HIV patients will have obstructive CAD), our model can avoid unnecessary coronary CTA in low (pretest probability <15%) and moderate risk ($15\% \leq \text{pretest probability} \leq 70\%$) as much as possible (11). Given that all features used in our Model 1—such as age, lipid profiles, blood pressure, ART duration, and CD4⁺ counts—are routinely collected in HIV outpatient settings, this model is well-suited for clinical application. It may serve as a reference framework for developing clinical decision support systems (CDSS) to assist clinicians in pre-test risk stratification of obstructive CAD among asymptomatic PLWH (39). In addition, for patients with renal abnormalities or other contraindications to coronary CTA, the model can also be used as a reference for testing obstructive CAD to assist clinicians in evaluation and

diagnosis and may help guide patient management. However, prospective studies with external validation and multi-center cohorts are needed to support its clinical integration and to evaluate its impact on clinical practice.

There are several limitations inherent to the present study. Firstly, due to the particular characteristics of the PLWH population, no external validation was conducted in an independent cohort in the current investigation. Although we have improved the algorithm, the actual clinical implementation and popularization of the model still need to be explored in further detail. Secondly, assessment of significant coronary stenosis was performed by coronary CTA without the use of invasive coronary angiography, although previous articles have demonstrated the high diagnostic value of coronary CTA for identifying coronary stenosis (40).

Conclusion

The random forest model with the inclusion of HIV-specific metrics showed better performance and best fit compared with the traditional models. The machine learning model better predicted the likelihood of obstructive CAD in low to moderate risk PLWH and may be helpful for clinical decision-making.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the First Hospital of China Medical University Technology ethics committee (AF-SOP-07-1.1-01). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

HuL: Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. HD: Conceptualization, Data curation, Investigation, Resources, Validation, Writing – review & editing. YZ: Data curation, Methodology, Writing – original draft. YL: Data curation, Investigation, Visualization, Writing – original draft. YY: Data curation, Methodology, Software, Validation, Writing – original draft. ZG: Data curation, Investigation, Writing – original draft. JZ: Data curation, Investigation, Writing – original draft. HH: Methodology, Writing – review & editing. HaL: Data curation, Methodology, Writing – original draft. XP: Conceptualization, Methodology, Writing – original draft. ZT: Data curation, Investigation, Writing – original draft. XL: Data curation, Methodology, Resources, Writing – review & editing. HS: Conceptualization, Data curation, Project administration, Resources, Writing – review & editing. TL:

Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. National Natural Science Foundation of China (No. 82171918); National Natural Science Foundation of China (No. 81871435); the Projects of Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences (2019PT320019).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor YW declared a shared parent affiliation with the author(s) HuL, HD, YZ, YL, ZG, JZ, HH, XP, ZT, XL, HS, and TL at the time of the review.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Bekker LG, Beyrer C, Mgodi N, Lewin SR, Delany-Moretlwe S, Taiwo B, et al. HIV infection. *Nat Rev Dis Primers*. (2023) 9:1–21. doi: 10.1038/s41572-023-00452-3
2. Cai CW, Sereti I. Residual immune dysfunction under antiretroviral therapy. *Semin Immunol*. (2021) 51:101471. doi: 10.1016/j.smim.2021.101471
3. Menéndez-Arias L, Delgado R. Update and latest advances in antiretroviral therapy. *Trends Pharmacol Sci*. (2022) 43:16–29. doi: 10.1016/j.tips.2021.10.004
4. Urina-Jassir M, Patiño-Aldana AF, Herrera-Parra LJ, Hernández Vargas JA, Trujillo-Cáceres SJ, Valbuena-García AM, et al. Factors associated with coronary artery disease among people living with human immunodeficiency virus: results from the Colombian HIV/AIDS registry. *Int J Cardiol Cardiovasc Risk Prev*. (2023) 18:200205. doi: 10.1016/j.ijcrp.2023.200205
5. Strategies for Management of Antiretroviral Therapy (SMART) Study Group (El-Sadr WM, Lundgren J, Neaton JD, Gordin F, Abrams D, Arduino RC, et al.) CD4+ count-guided interruption of antiretroviral treatment. *N Engl J Med*. (2006) 355:2283–96. doi: 10.1056/NEJMoa062360
6. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, et al. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. *Circulation*. (2023) 147:e93–e621. doi: 10.1161/CIR.0000000000001123
7. Stocker TJ, Deseive S, Leipsic J, Hadamitzky M, Chen MY, Rubinshtein R, et al. Reduction in radiation exposure in cardiovascular computed tomography imaging: results from the PROspective multicenter registry on radiation dose estimates of cardiac CT angiography in daily practice in 2017 (PROTECTION VI). *Eur Heart J*. (2018) 39:3715–3723. doi: 10.1093/eurheartj/ehy546
8. Douglas PS, Hoffmann U, Patel MR, Mark DB, Al-Khalidi HR, Cavanaugh B, et al. Outcomes of anatomical versus functional testing for coronary artery disease. *N Engl J Med*. (2015) 372:1291–300. doi: 10.1056/NEJMoa1415516

9. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur. Heart J.* (2020) 41:407–477. doi: 10.1093/eurheartj/ehz425
10. Genders TSS, Steyerberg EW, Hunink MGM, Nieman K, Galema TW, Mollet NR, et al. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. *BMJ.* (2012) 344:e3485. doi: 10.1136/bmj.e3485
11. Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, Budaj A, et al. ESC guidelines on the management of stable coronary artery disease: the task force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur Heart J.* (2013) 34:2949–3003. doi: 10.1093/eurheartj/ehz296
12. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med.* (2023) 388:1201–8. doi: 10.1056/NEJMra2302038
13. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* (2018) 2:719–31. doi: 10.1038/s41551-018-0305-z
14. Peng X, Ding H, Huo H, Zheng Y, Zhou J, Li H, et al. Cardiac MRI -based assessment of myocardial injury in asymptomatic people living with human immunodeficiency virus: correlation with nadir CD4 count. *Magn Reson Imaging.* (2023) 58:1815–23. doi: 10.1002/jmri.28699
15. Austen WG, Edwards JE, Frye RL, Gensini GG, Gott VL, Griffith LS, et al. A reporting system on patients evaluated for coronary artery disease. Report of the ad hoc Committee for Grading of coronary artery disease, council on cardiovascular surgery, American Heart Association. *Circulation.* (1975) 51:5–40. doi: 10.1161/01.cir.51.4.5
16. Mancini GBJ, Leipsic J, Budoff MJ, Hague CJ, Min JK, Stevens SR, et al. CT angiography followed by invasive angiography in patients with moderate or severe Ischemia-insights from the ISCHEMIA trial. *JACC Cardiovasc Imaging.* (2021) 14:1384–93. doi: 10.1016/j.jcmg.2020.11.012
17. Gatti M, Gallone G, Poggi V, Bruno F, Serafini A, Depaoli A, et al. Diagnostic accuracy of coronary computed tomography angiography for the evaluation of obstructive coronary artery disease in patients referred for transcatheter aortic valve implantation: a systematic review and meta-analysis. *Eur Radiol.* (2022) 32:5189–200. doi: 10.1007/s00330-022-08603-y
18. Collet J-P, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, et al. 2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur Heart J.* (2021) 42:1289–367. doi: 10.1093/eurheartj/ehaa575
19. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol.* (1990) 15:827–32. doi: 10.1016/0735-1097(90)90282-t
20. Knuuti J. ESC guidelines for the diagnosis and management of chronic coronary syndromes the task force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Russ J Cardiol.* (2019) 25:119–80. doi: 10.15829/1560-4071-2020-2-3757
21. Task Force Members; Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur. Heart J.* (2013) 34. doi: 10.1093/eurheartj/ehz296
22. AlAref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J.* (2019) 40:1975–86. doi: 10.1093/eurheartj/ehy404
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
24. Bittencourt MS, Hultén E, Polonsky TS, Hoffman U, Nasir K, Abbasa S, et al. European society of cardiology-recommended coronary artery disease consortium pretest probability scores more accurately predict obstructive coronary disease and cardiovascular events than the diamond and Forrester score: the partners registry. *Circulation.* (2016) 134:201–11. doi: 10.1161/CIRCULATIONAHA.116.023396
25. AlAref SJ, Maliakal G, Singh G, Van Rosendaal AR, Ma X, Xu Z, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *Eur Heart J.* (2020) 41:359–67. doi: 10.1093/eurheartj/ehz565
26. Winther S, Nissen L, Westra J, Schmidt SE, Bouteldja N, Knudsen LL, et al. Pre-test probability prediction in patients with a low to intermediate probability of coronary artery disease: a prospective study with a fractional flow reserve endpoint. *Eur. Heart J.* (2019) 20:1208–18. doi: 10.1093/ehjci/jez058
27. Ramrakhiani NS, Chen VL, Le M, Yeo YH, Barnett SD, Waljee AK, et al. Optimizing hepatitis B virus screening in the United States using a simple demographics-based model. *Hepatology.* (2022) 75:430–7. doi: 10.1002/hep.32142
28. Li J, Tian Y, Zhu Y, Zhou T, Li J, Ding K, et al. A multicenter random forest model for effective prognosis prediction in collaborative clinical research network. *Artif Intell Med.* (2020) 103:101814. doi: 10.1016/j.artmed.2020.101814
29. Freiberg MS, Chang C-CH, Kuller LH, Skanderson M, Lowy E, Kraemer KL, et al. HIV infection and the risk of acute myocardial infarction. *JAMA Intern Med.* (2013) 173:614–22. doi: 10.1001/jamainternmed.2013.3728
30. Triant VA, Lee H, Hadigan C, Grinspoon SK. Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease. *J Clin Endocrinol Metabol.* (2007) 92:2506–12. doi: 10.1210/jc.2006-2190
31. Post WS, Budoff M, Kingsley L, Palella FJ, Witt MD, Li X, et al. Associations between HIV infection and subclinical coronary atherosclerosis. *Ann Intern Med.* (2014) 160:458–67. doi: 10.7326/M13-1754
32. Boccara F, Lang S, Meuleman C, Ederhy S, Mary-Krause M, Costagliola D, et al. HIV and coronary heart disease. *J Am Coll Cardiol.* (2013) 61:511–23. doi: 10.1016/j.jacc.2012.06.063
33. Smit M, Brinkman K, Geerlings S, Smit C, Thyagarajan K, Sighem AV, et al. Future challenges for clinical care of an ageing population infected with HIV: a modelling study. *Lancet Infect Dis.* (2015) 15:810–8. doi: 10.1016/S1473-3099(15)00056-0
34. Lichtenstein KA, Armon C, Buchacz K, Chmiel JS, Buckner K, Tedaldi EM, et al. Low CD4⁺ T cell count is a risk factor for cardiovascular disease events in the HIV outpatient study. *Clin Infect Dis.* (2010) 51:435–47. doi: 10.1086/655144
35. Serrano-Villar S, Sainz T, Lee SA, Hunt PW, Sinclair E, Shacklett BL, et al. HIV-infected individuals with low CD4/CD8 ratio despite effective antiretroviral therapy exhibit altered T cell subsets, heightened CD8⁺ T cell activation, and increased risk of non-AIDS morbidity and mortality. *PLoS Pathog.* (2014) 10:e1004078. doi: 10.1371/journal.ppat.1004078
36. Serrano-Villar S, Pérez-Eliás MJ, Dronda F, Casado JL, Moreno A, Royuela A, et al. Increased risk of serious non-AIDS-related events in HIV-infected subjects on antiretroviral therapy associated with a low CD4/CD8 ratio. *PLoS One.* (2014) 9:e85798. doi: 10.1371/journal.pone.0085798
37. Tarr PE, Ledergerber B, Calmy A, Doco-Lecompte T, Marzel A, Weber R, et al. Subclinical coronary artery disease in Swiss HIV-positive and HIV-negative persons. *Eur Heart J.* (2018) 39:2147–54. doi: 10.1093/eurheartj/ehy163
38. Brockow K, Wurpts G, Trautmann A, Pfütznern W, Treudler R, Bircher AJ, et al. Guideline for allergological diagnosis of drug hypersensitivity reactions. *Allergol Select.* (2023) 7:122–39. doi: 10.5414/ALX02422E
39. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med.* (2020) 3:17. doi: 10.1038/s41746-020-0221-y
40. Leschka S, Alkadhi H, Plass A, Desbiolles L, Grünenfelder J, Marinček B, et al. Accuracy of MSCT coronary angiography with 64-slice technology: first experience. *Eur Heart J.* (2005) 26:1482–7. doi: 10.1093/eurheartj/ehi261