

Research Article

Development of Machine Learning Tools for Predicting Coronary Artery Disease in the Chinese Population

Tiexu Zhang ¹, Shengming Huang ^{2,3}, Pengfei Xie ¹, Xiaoming Li ^{4,5}, Yingxia Pan ^{4,5},
Yue Xu ^{4,5}, Peng Han ^{4,5}, Feifei Ding ^{4,5}, Jiangman Zhao ^{4,5}, and Hui Tang ^{4,5}

¹Department of Cardiovascular Medicine, The First People's Hospital of Pingdingshan, Pingdingshan 467000, China

²Department of Internal Medicine, Luohe Central Hospital, Luohe, 462000 Henan, China

³Health Management Center, Luohe Central Hospital, Luohe 462000, China

⁴Shanghai Biotecan Pharmaceuticals Co., Ltd, Shanghai 201204, China

⁵Shanghai Zhangjiang Institute of Medical Innovation, Shanghai 201204, China

Correspondence should be addressed to Tiexu Zhang; zhangtiexu@163.com, Jiangman Zhao; zhaojiangman86@163.com, and Hui Tang; tang11_23@126.com

Received 14 July 2022; Revised 9 October 2022; Accepted 1 November 2022; Published 17 November 2022

Academic Editor: Gaetano Isola

Copyright © 2022 Tiexu Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. Coronary artery disease (CAD) is one of the major cardiovascular diseases and the leading cause of death globally. Blood lipid profile is associated with CAD early risk. Therefore, we aim to establish machine learning models utilizing blood lipid profile to predict CAD risk. **Methods.** In this study, 193 non-CAD controls and 2001 newly-diagnosed CAD patients (1647 CAD patients who received lipid-lowering therapy and 354 who did not) were recruited. Clinical data and the result of routine blood lipids tests were collected. Moreover, low-density lipoprotein cholesterol (LDL-C) subfractions (LDLC-1 to LDLC-7) were classified and quantified using the Lipoprint system. Six predictive models (k-nearest neighbor classifier (KNN), logistic regression (LR), support vector machine (SVM), decision tree (DT), multilayer perceptron (MLP), and extreme gradient boosting (XGBoost)) were established and evaluated by the confusion matrix, area under the receiver operating characteristic (ROC) curve (AUC), recall (sensitivity), accuracy, precision, and F1 score. The selected features were analyzed and ranked. **Results.** While predicting the CAD development risk of the CAD patients without lipid-lowering therapy in the test set, all models obtained AUC values above 0.94, and the accuracy, precision, recall, and F1 score were above 0.84, 0.85, 0.92, and 0.88, respectively. While predicting the CAD development risk of all CAD patients in the test set, all models obtained AUC values above 0.91, and the accuracy, precision, recall, and F1 score were above 0.87, 0.94, 0.87, and 0.92, respectively. Importantly, small dense LDL-C (sdLDL-C) and LDLC-4 play pivotal roles in predicting CAD risk. **Conclusions.** In the present study, machine learning tools combining both clinical data and blood lipid profile showed excellent overall predictive power. It suggests that machine learning tools are suitable for predicting the risk of CAD development in the near future.

1. Introduction

Coronary artery disease (CAD) is a cardiovascular disease (CVD) which has been found to be the leading cause of mortality worldwide [1] and caused by atherosclerosis, which can be manifested by typical symptoms such as stable angina, unstable angina, myocardial infarction (MI), or sudden cardiac death without any preceding symptoms [2]. So

far, coronary angiography is the gold method for CAD diagnosis, but it is an invasive and unpractical for universal screening [3]. Hence, finding cost-effective methods to predict CAD are a major challenge in public health.

Being a complex disease, CAD is caused by genetic and environmental factors as well as the interactions between these factors [4]. The well-known risk factors for CAD development include hypertension, dyslipidemia, older age,

diabetes mellitus, overweight, and smoking [5, 6]. Furthermore, the prevalence of CAD varies greatly according to the geographical locations, ethnicity, and gender [7].

Besides that, a previous research study indicated that low-density lipoprotein cholesterol (LDL-C) is a primary risk factor for CVD [8], and lowering LDL-C levels with medications have been proved to be effective for primary and secondary prevention [9]. LDL-C is composed of heterogeneous particles with different density and size, which could be classified into two subgroups including small dense LDL-C (sdLDL-C) and large buoyant LDL-C (lbLDL-C) [10]. Moreover, LDL-C was divided into 7 subfractions (LDLC-1 to LDLC-7), of which LDLC-1 and LDLC-2 belong to lbLDL-C, while LDLC-3 to LDLC-7 is defined as sdLDL-C, according to their density and size [10]. Over the past few years, many preventive and therapeutic methods have substantially improved the prognosis of patients with CAD or other CVD [2, 11]. However, the risk of such diseases remains high, and their progression could be halted only in a few patients by using drugs including aspirin, statins, and β -blockers [12]. When lowering LDL-C levels to optimal levels, the risk of cardiovascular events still exists [13]. The Atherosclerosis Risk in Communities (ARIC) study showed that sdLDL-C levels may partly account for this residual risk. Since sdLDL-C particles contain less cholesterol and are smaller, increased sdLDL-C levels represent an increase in the amount of atherogenic LDL particles, which LDL-C levels may not represent [14]. In addition, a previous research study has shown that sdLDL-C is considered as an important biomarker for predicting CVD [15]. It is reported that sdLDL-C has stronger transfer ability moving from the vessel lumen into the subintimal space [16], weaker binding affinity to LDL-C receptors [17], and longer plasma residence time [18]. Moreover, Srisawasdi et al. reported that the ratio of sdLDL-C/lbLDL-C was a potential biomarker for assessing lipid metabolic status in patients with metabolic syndrome [19]. Numerous epidemiologic studies and randomized clinical trials have suggested that elevated LDL-C is a major cause of CAD and the target to be controlled to reduce atherosclerotic cardiovascular disease risk [6, 20, 21]. However, a large proportion of atherosclerosis and CAD patients have normal range of blood LDL-C level. Recent large cohort studies have demonstrated that using a simple homogeneous sdLDL-C assay can predict the cardiovascular risk regardless of LDL-C level [15, 22–24]. These findings suggest that total LDL-C level cannot completely represent its biological effect and cannot fully represent blood lipid levels. Although many studies have shown that 10-year Framingham risk score and atherosclerotic CVD risk score were developed based on hypertension, LDL-C, HDL-C, TG, TC, age, smoking, and diabetes risk factors and used to predict CVD risk [25, 26], these prediction models have been reported that they have limitations to estimate future CVD risk [27, 28]. Hence, it is urgent to explore more potential risk factors. Therefore, more attention should be paid to explore the relationship between LDL-C subfractions and CAD risk, which may help to elucidate the differences among patients with CAD, and to establish the early warning to assess the risk of CAD.

In this study, we recruited 193 non-CAD controls and 2001 newly CAD patients including 1647 CAD patients who received lipid-lowering therapy and 354 who did not, collected their clinical features, and measured the concentration of blood lipid profile, including total cholesterol (TC), triglyceride (TG), high density lipoprotein cholesterol (HDL-C), LDL-C, and LDL-C subfractions. We aim to establish and evaluate six machine learning tools combining clinical features and blood lipid profile could predict the risk of CAD patients who did not receive lipid-lowering therapy or all CAD patients including 1647 CAD patients who received lipid-lowering therapy and 354 who did not.

2. Materials and Methods

2.1. Study Population. A total of 2001 newly diagnosed CAD consecutive patients including 1647 CAD patients who received lipid-lowering therapy and 354 who did not, as well as 193 non-CAD controls were recruited from the First People's Hospital of Pingdingshan and Luohe Central Hospital, from July 2018 to October 2019. The inclusion criteria for participants were as follows: (i) CAD patients were diagnosed by coronary angiography, which is defined as coronary artery stenosis $\geq 50\%$ in at least one main vessel or its major branches as described [29]; (ii) the non-CAD controls were diagnosed by coronary angiography without any luminal stenosis or plaque in main vessels and branches; (iii) the age of all participants > 18 years. The exclusion criteria were as follows: (i) patients who had prior CAD or revascularization (percutaneous or surgical) [30]; (ii) participants who do not understand this research study [3]; (iii) participants who had severe medical disease, such as liver or kidney disease, thyroid disease, and malignant diseases [29], as well as immune-related sickness, nephropathic diseases, and respiratory diseases and also physiological conditions related to immune responses such as pregnancy [31]. The exclusion criteria for the non-CAD controls were the same as what mentioned above. Clinical information of all participants was collected including age, gender, body mass index (BMI), smoking and drinking status, and other diseases history such as hypertension and diabetes mellitus.

Informed consent was obtained from each participant included in the study, and the study protocol conforms to the ethical guidelines of the latest version of Declaration of Helsinki, and the study protocol has been approved by Ethical Committee of the First People's Hospital of Pingdingshan and Luohe Central Hospital.

2.2. Blood Lipid Profile Detection. Blood samples were collected by serum separator tube and anticoagulant tube. Plasma was separated immediately after collection by $800 \times g$ centrifugation for 10 min at 4°C . TC, TG, HDL-C, LDL-C, and glucose were measured in the Department of Clinical Laboratory. The reference ranges of TC, TG, HDL-C, and LDL-C were defined by 2016 Chinese guideline for the management of dyslipidemia in adults [32].

LDL-C subfractions were classified and quantified by LDL subfractions kit of Shanghai Biotecan Pharmaceuticals Co., Ltd. Briefly, the plasma mixed with liquid loading gel

was added to the top of precast 3% polyacrylamide gel tubes. After 30 min of photopolymerization at room temperature, samples were electrophoresed in electrophoresis apparatus (Shanghai Biotecan Pharmaceuticals Co., Ltd.) for 70 min (3 mA/tub). Then, the densitometry was determined by Gel Scanner (Hunan Biotecan Medical Device Co., Ltd.). Last, LDL-C was separated into 7 subfractions (LDLC-1 to LDLC-7) as previously described [10].

2.3. Machine Learning Tools In Construction. In the present study, six machine learning tools were established to predict CAD risk, including k-nearest neighbor classifier (KNN) model, logistic regression (LR) model, support vector machine (SVM) model, decision tree (DT) model, multilayer perceptron (MLP) model, and extreme gradient boosting (XGBoost) model. Clinical characteristics and blood lipid profile are fit in the models. These models were tuned using a set of parameters, which were adjusted to obtain the average performance index. The tuning parameters of the six prediction models are listed for the optimization of the equations (Table S1). Python (version 3.8) was used as the basic language in the whole model, and NumPy, pandas, sklearn, XGBoost, and Matplotlib libraries were used to process the data and establish the models.

When to predict the CAD development risk in patients without lipid-lowering therapy, a total of 547 participants (193 non-CAD controls and 354 CAD patients) were randomly allocated into a training set (80%) and a test set (20%). When to predict the CAD development risk in all CAD patients, a total of 2194 participants (193 non-CAD controls and 2001 CAD patients) were randomly allocated into a training set (80%) and a test set (20%). In the training set, StratifiedKFold ($k = 5$) was used, and various parameter combinations were exhausted using grid search. For each model, the confusion matrix, area under the receiver operating characteristic (ROC) curve (AUC), accuracy, recall (sensitivity), precision, and F1 score were used to evaluate and compare the comprehensive performance of feature selection [33]. AUC is the main metric in evaluating binary classifiers and shows the true positive rate against the false positive rate [33]. Precision and recall are excellent metrics for capturing the aspects of model performance [34]. The F1 score takes the geometric mean of precision and recall [35]. In addition, the feature score (F score) rankings were measured by the total_gain metric in XGBoost model [3].

2.4. Statistical Analysis. Statistical analyses were performed using GraphPad Prism (version 6.0; GraphPad Software, Inc.) or SPSS 19.0 (IBM, NY, USA) and R 3.5.1 software. Categorical variables were presented by numbers or proportions, and differences in distribution between two groups were analyzed by chi-squared test. Continuous variables were presented using median with interquartile range (IQR) because they are non-Gaussian distributions data [36]. Nonparametric Mann-Whitney U test was used to analyze the difference between two groups or Kruskal-Wallis H test followed by Dunn's post hoc test that was used to analyze the difference among three groups. Correlation analysis among CAD development, clinical features, and

blood lipid profile was conducted by the Pearson correlation method. A $P < 0.05$ was considered statistically significant.

3. Results

3.1. Comparison of Clinical Characteristics between the Non-CAD Controls and CAD Patients. The clinical characteristics of all participants (193 non-CAD controls and 2001 CAD patients) were collected in Table 1. The CAD patients were older than the non-CAD controls. The median BMI of CAD patients was significantly higher than that of the non-CAD controls. Moreover, a noticeably higher prevalence of drinking, hypertension, and diabetes mellitus was observed in CAD patients, comparing to non-CAD controls (Table 1).

3.2. Comparison of the Blood Lipid Profile Levels among the Non-CAD Controls, CAD Patients Who Received Lipid-Lowering Therapy, and those Who Did Not. We detected the levels of TC, TG, HDL-C, and LDL-C and compared among 193 non-CAD controls (control group), 1647 CAD patients who received lipid-lowering therapy (lipid-lowering therapy group), and 354 CAD patients did not receive any lipid-lowering therapy (non-lipid-lowering therapy group). The level of TC is significantly higher in the two groups of CAD patients than that in the control group. The level of TG is the highest in the lipid-lowering therapy group than that in the other two groups (Figures 1(a) and 1(b)). Moreover, HDL-C is the highest in the control group than in the two groups of CAD patients, while LDL-C was the lowest in the control group (Figures 1(c) and 1(d)). However, there are no HDL-C and LDL-C differences between the two groups of CAD patients.

In addition, we also compared the levels of LDL-C subfractions (LDLC-1 to LDLC-7) among the three groups. The level of LDLC-1 was significantly lower, while LDLC-2 was noticeably higher in the two groups of CAD patients than that in the control group. The total lLDL-C level was only higher in the control group than in the other two groups (Figures 2(a)–2(c)). The concentrations of LDLC-3, LDLC-4, LDLC-5, and sdLDL-C are significantly higher in both two groups of CAD patients than that in the control group. Moreover, the concentrations of LDLC-4, LDLC-5, and sdLDL-C are significantly higher in the lipid-lowering therapy than that in the non-lipid-lowering therapy group. However, both LDLC-6 and LDLC-7 showed no significant differences among the three groups (Figures 2(d)–2(i)).

3.3. The Abnormal Rates of Blood Lipid Profile among in the Non-CAD Controls and the CAD Patients Who Received Lipid-Lowering Therapy and those Who Did Not. Besides that, we found that the abnormal rates of TC, TG, HDL-C, and LDL-C were 11.92%, 15.03%, 27.98%, and 4.66% in the control group; 20.34%, 37.57%, 34.46%, and 15.25% in the non-lipid-lowering therapy group; 19.31%, 47.42%, 35.03%, and 12.26% in the lipid-lowering therapy group (Table 2), respectively. Moreover, we further investigated the abnormal rates of LDL-C subfractions separately. Surprisingly, the abnormal rates of LDLC-3, LDLC-4, and

TABLE 1: The clinical characteristics of 193 non-CAD controls and 2001 CAD patients.

Characteristics	Non-CAD controls (n = 193)	CAD patients (n = 2001)	p value
Age (years) (Median with IRQ)	52 (43-68)	64 (54-72.5)	<0.0001
Gender			0.51
(i) Male	99	1076	
(ii) Female	94	925	
BMI (kg/m ²) (median with IRQ)	22.86 (20.83-26.04)	24.69 (23.40-27.24)	<0.0001
Smoking			0.065
(i) Yes	24	354	
(ii) No	169	1647	
Drinking			0.008
(i) Yes	20	358	
(ii) No	173	1643	
Hypertension			<0.0001
(i) Yes	43	1481	
(ii) No	150	520	
Diabetes mellitus			<0.0001
(i) Yes	43	775	
(ii) No	150	1226	
Lipid-lowering therapy			/
(i) Yes	0	1647	
(ii) No	193	354	

CAD: coronary artery disease; IRQ: interquartile range; BMI: body mass index; * $p < 0.05$ was considered to be significant.

LDLC-5 were 80.79%, 76.27%, and 24.29% in the non-lipid-lowering therapy group and 84.52%, 84.46%, and 32.67% in the lipid-lowering therapy group, respectively. On the contrary, the abnormal rates of LDLC-3 (3.11%), LDLC-4 (1.04%), and LDLC-5 (0.52%) were very low in the control group (Table 2).

3.4. Comparison of the sdLDL-C Subfractions among the Non-CAD Controls, the CAD Patients Who Received Lipid-Lowering Therapy, and those Who Did Not. According to the above results, we found the very high abnormal rates of LDLC-3 and LDLC-4 in the lipid-lowering therapy and non-lipid-lowering therapy groups. However, it remains unknown that whether the abnormal rates of LDLC-3 and LDLC-4 in CAD patients with normal LDL levels are high or not. Thus, we divided LDL-C normal and LDL-C high subgroups among the three groups separately and detected the abnormal rates of sbLDL-C subfractions in the two subgroups. Surprisingly, the abnormal rates of LDLC-3 and LDLC-4 were still high in all CAD patients with normal LDL-C levels. To be specific, the abnormal rates of LDLC-3 and LDLC-4 were 82.98% and 84.43%, respectively, in the lipid-lowering therapy group, while the abnormal rates of LDLC-3 and LDLC-4 were 78.33% and 75.67%, respectively, in the non-lipid-lowering therapy group. However, the abnormal rates of these two LDL-C subfractions remained low in the non-CAD controls with normal LDL-C levels (Table 3). These results indicated that LDLC-3 and

LDLC-4 were the main components in the sbLDL-C subfractions and may play an important role in CAD development.

3.5. Correlation Analysis among CAD Risk, Clinical Characteristics, and Blood Lipid Profile. In order to analyze the correlation among clinical data, blood lipid profile, and CAD risk, Pearson correlation analysis was employed to evaluate the correlation among them (Figure 3). Since sex and smoking had no significant differences between the non-CAD controls and CAD patients (Table 1), and both LDLC-6 and LDLC-7 were not noticeably expressed among the three groups (Figures 2(g) and 2(h)), sex, smoking, LDLC-6, and LDLC-7 were excluded in the Pearson correlation analysis. We found that age ($r = 0.2$, $p < 0.001$), hypertension ($r = 0.32$, $p < 0.001$), LDLC-3 ($r = 0.4$, $p < 0.001$), LDLC-4 ($r = 0.27$, $p < 0.001$), and sdLDL-C ($r = 0.36$, $p < 0.001$) were significantly positively correlated with CAD, while LDLC-1 ($r = -0.21$, $p < 0.001$) was significantly negatively correlated with CAD. Moreover, TC was significantly positively correlated with HDL-C ($r = 0.3$, $p < 0.001$), LDL-C ($r = 0.89$, $p < 0.001$), LDLC-1 ($r = 0.38$, $p < 0.001$), LDLC-2 ($r = 0.41$, $p < 0.001$), LDLC-3 ($r = 0.41$, $p < 0.001$), LDLC-4 ($r = 0.35$, $p < 0.001$), sdLDL-C ($r = 0.42$, $p < 0.001$), and lbLDL-C ($r = 0.38$, $p < 0.001$). HDL-C had strongly positive correlation with LDLC-1 ($r = 0.4$, $p < 0.001$), LDLC-2 ($r = 0.28$, $p < 0.001$), and lbLDL-C ($r = 0.38$, $p < 0.001$). Meanwhile, LDL-C had significantly positive

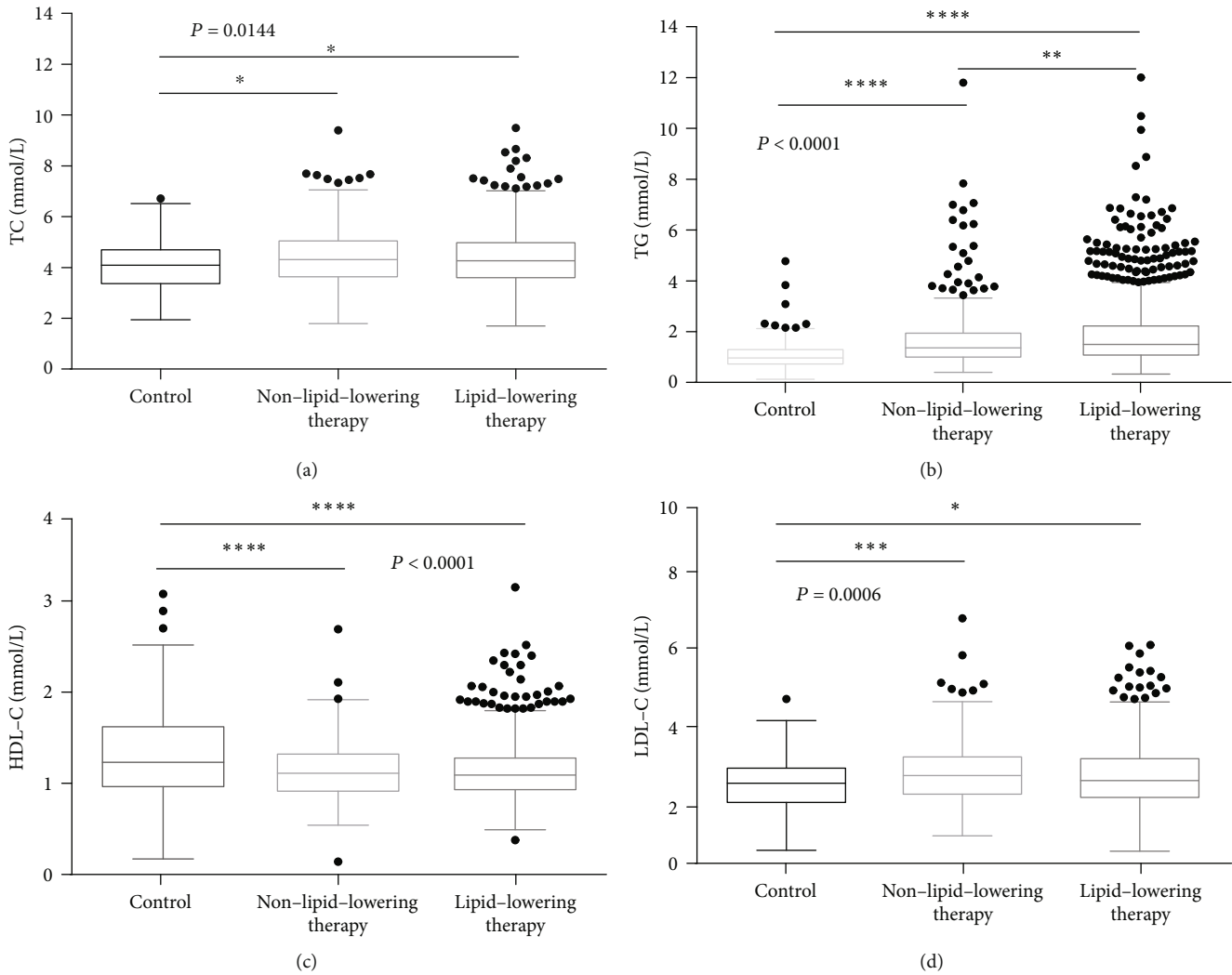


FIGURE 1: Comparison of blood lipids level among non-CAD controls and CAD patients who received lipid-lowering therapy or those who did not. The different expressions of (a) TC, (b) TG, (c) HDL-C, and (d) LDL-C among three groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

correlation with LDL-C1 ($r = 0.44$, $p < 0.001$), LDL-C2 ($r = 0.49$, $p < 0.001$), LDL-C3 ($r = 0.45$, $p < 0.001$), and LDL-C4 ($r = 0.31$, $p < 0.001$), as well as sdLDL-C ($r = 0.451$, $p < 0.001$) and lbLDL-C ($r = 0.53$, $p < 0.001$). In addition, lbLDL-C had strongly positive correlation with LDL-C1 ($r = 0.89$, $p < 0.001$) and LDL-C2 ($r = 0.87$, $p < 0.001$), while lbLDL-C was significantly negatively correlated with LDL-C4 ($r = -0.26$, $p < 0.001$) and LDL-C5 ($r = -0.25$, $p < 0.001$). Importantly, sdLDL-C had strongly positive correlation with LDL-C3 ($r = 0.86$, $p < 0.001$), LDL-C4 ($r = 0.92$, $p < 0.001$), and LDL-C5 ($r = 0.61$, $p < 0.001$), while sdLDL-C was significantly negatively correlated with LDL-C1 ($r = -0.3$, $p < 0.001$) (Figure 3).

3.6. Establish and Compare Six Machine Learning Models to Predict CAD Development. On the one hand, in order to predict the CAD risk of patients who did not receive lipid-lowering therapy, we firstly used XGBoost model to analyze the importance of features including 7 clinical features (age, sex, BMI, smoking, drinking, hypertension, and diabetes

mellitus) and 14 blood lipid profile (TC, TG, HDL-C, LDL-C, sdLDL-C, lbLDL-C, LDL-C1 to LDL-C7, and sdLDL-C/lbLDL-C), and the feature score (F score) rankings were measured by the total_gain metric in XGBoost. However, only 16 factors (age, sex, BMI, smoking, drinking, hypertension, TC, TG, HDL-C, LDL-C, sdLDL-C, lbLDL-C, LDL-C1, LDL-C2, LDL-C4, and sdLDL-C/lbLDL-C) were obtained in the rankings (Figure 4(a)). Among them, sdLDL-C, LDL-C4, and hypertension ranked top 3 in the feature importance rankings, while sex, TG, and lbLDL-C ranked lower (Figure 4(a)). Therefore, these 16 factors were enrolled into the six machine learning tools.

After that, a total of 547 participants (193 non-CAD controls and 354 CAD patients without receiving lipid-lowering therapy) were randomly allocated into a training set (80%) and a test set (20%). StratifiedKFold ($k = 5$) was used in the training set. After fitting in the training set, each model is evaluated by the test set. For each model, the evaluation indicators used were the confusion matrix, AUC, recall (sensitivity), precision, accuracy, and F1 score. The

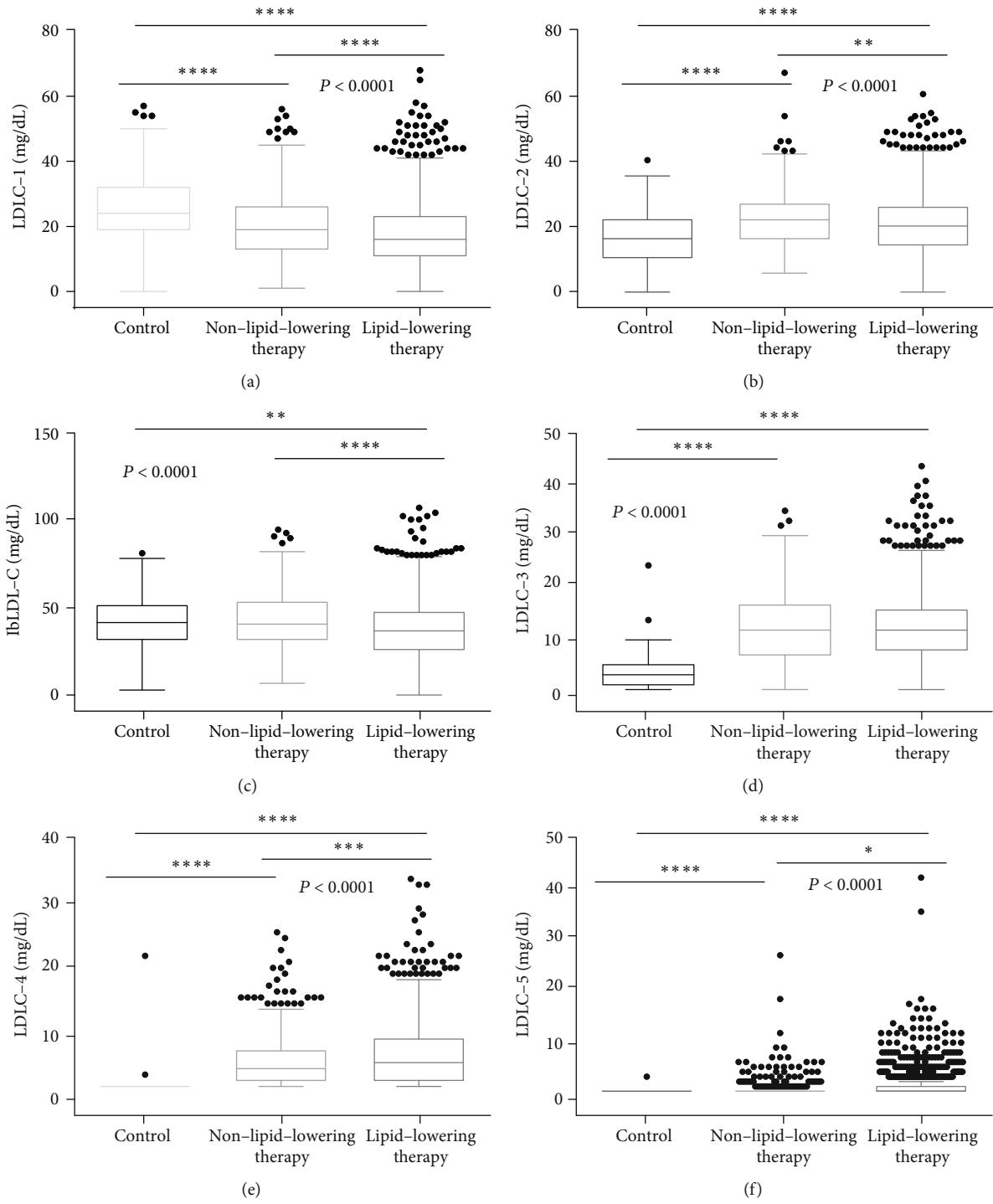


FIGURE 2: Continued.

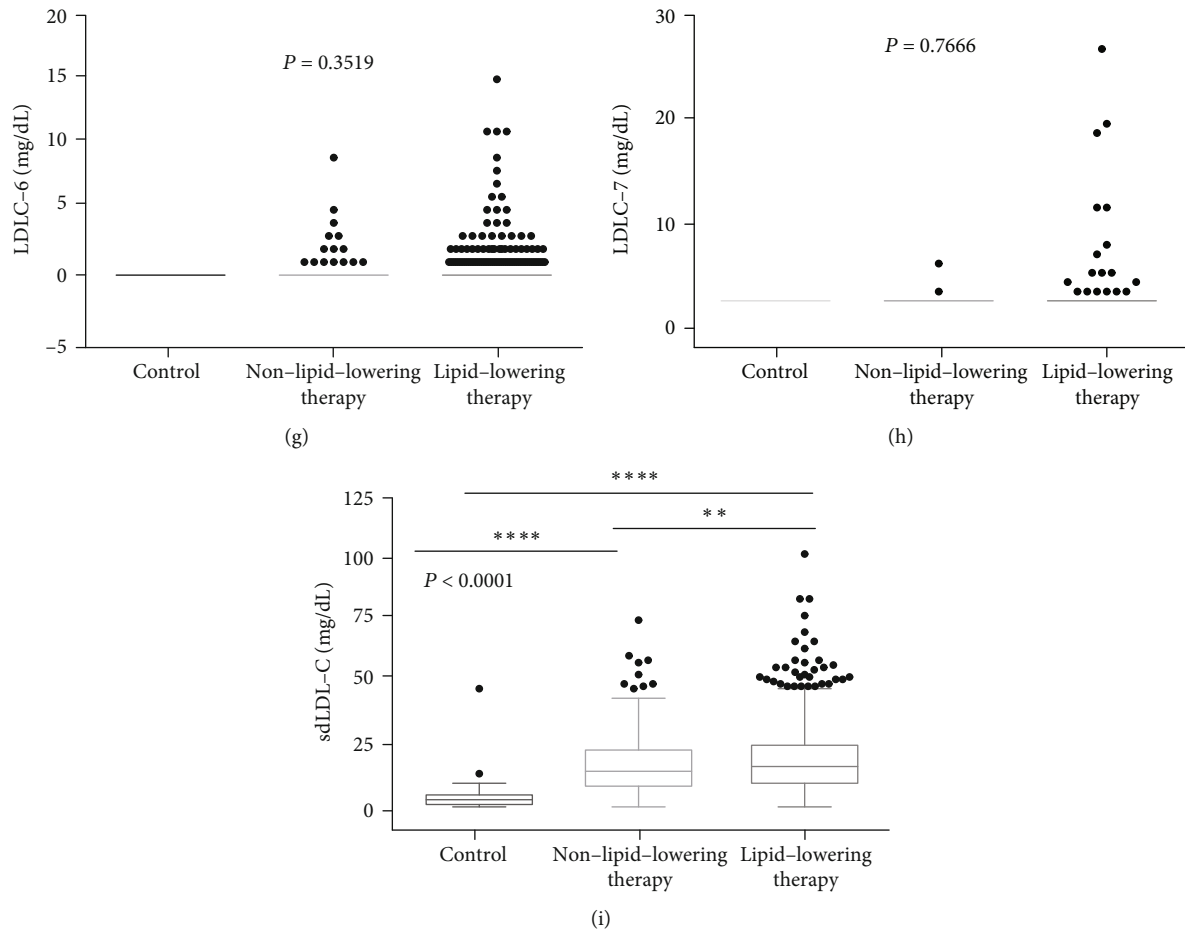


FIGURE 2: Comparison of LDL-C subfractions levels among non-CAD controls and CAD patients who received lipid-lowering therapy or those who did not. The different expressions of (a) LDL-C-1, (b) LDL-C-2, (c) lbLDL-C, (d) LDL-C-3, (e) LDL-C-4, (f) LDL-C-5, (g) LDL-C-6, (h) LDL-C-7, and (i) sdLDL-C among three groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

ROC curve is widely used to validate the performance of prediction models, and the average AUC and 95% CI are shown in Figures 4(b) and 4(c). In the training set, all models had AUC values above 0.90 (Figure 4(b)), and the accuracy, precision, recall, and F1 score were above 0.83, 0.87, 0.85, and 0.86, respectively (Table 4). Among them, XGBoost model had the highest AUC (0.95), as well as the highest accuracy (0.90), precision (0.94), and F1 score (0.92). In addition, the MLP model also obtained the same highest F1 score (0.92) and the highest recall (0.91). Importantly, in the test set, all models obtained the AUC values above 0.94 (Figure 4(c)), and the accuracy, precision, recall, and F1 score were above 0.84, 0.85, 0.92 and 0.88, respectively (Table 5). Interestingly, XGBoost still obtained the highest AUC (0.98), accuracy (0.93), precision (0.93), recall (0.96), and F1 score (0.94). Meanwhile, the DT model obtained the same highest precision (0.93) and F1 score (0.94).

On the other hand, in order to verify whether the six machine learning tools combining clinical features and blood lipid profile could predict the risk of all CAD patients including 1647 CAD patients who received lipid-lowering therapy and 354 who did not, we enrolled all CAD patients

in the six predictive models. We also used the XGBoost model to analyze the importance of 21 features as above mentioned. Finally, only 16 factors obtained the importance ranking including age, BMI, smoking, drinking, hypertension, TC, TG, HDL-C, LDL-C, sdLDL-C, lbLDL-C, LDL-C-1, LDL-C-2, LDL-C-3, LDL-C-4, and sdLDL-C/lbLDL-C. Interestingly, sdLDL-C, LDL-C-4, and sdLDL-C/lbLDL-C ranked top 3 in the feature importance rankings, while LDL-C-2, LDL-C-3, and smoking ranked lower (Figure 5(a)). Thus, these 16 factors were enrolled into the six machine learning tools.

After that, a total of 2194 participants (193 non-CAD controls and 2001 CAD patients) were randomly allocated into a training set (80%) and a test set (20%). In the training set, all models obtained AUC values above 0.92 (Figure 5(b)), and the accuracy, precision, recall, and F1 score were above 0.89, 0.94, 0.89, and 0.93, respectively (Table 6). Among them, XGBoost model had the highest AUC (0.98), as well as the highest accuracy (0.95), recall (0.98), and F1 score (0.97). Furthermore, the LR, SVM, and DT model obtained the same highest precision (0.99). MLP showed the same highest accuracy (0.95) and F1 score (0.97). Importantly, in the test set, all models obtained the

TABLE 2: The abnormal rate of blood lipid levels among the non-CAD controls, the CAD patients who received lipid-lowering therapy, and those who did not.

Variables	Threshold	Control group (n = 193)	Non-lipid-lowering therapy group (n = 354)	Lipid-lowering therapy group (n = 1647)
		Abnormal rate	Abnormal rate	Abnormal rate
TC (mmol/L)	≥ 5.20	11.92% (23/193)	20.34% (72/354)	19.31% (318/1647)
TG (mmol/L)	≥ 1.7	15.03% (29/193)	37.57% (133/354)	47.42% (781/1647)
HDL-C (mmol/L)	< 1	27.98% (54/193)	34.46% (122/354)	35.03% (577/1647)
LDL-C (mmol/L)	≥ 3.4	4.66% (9/193)	15.25% (54/354)	12.26% (202/1647)
LDLC-1 (mg/dL)	/	/	/	/
LDLC-2 (mg/dL)	/	/	/	/
LDLC-3 (mg/dL)	> 6	3.11% (6/193)	80.79% (286/354)	84.52% (1392/1647)
LDLC-4 (mg/dL)	> 0	1.04% (2/193)	76.27% (270/354)	84.46% (1391/1647)
LDLC-5 (mg/dL)	> 0	0.52% (1/193)	24.29% (86/354)	32.67% (538/1647)
LDLC-6 (mg/dL)	> 0	0.00% (0/193)	4.24% (15/354)	5.34% (88/1647)
LDLC-7 (mg/dL)	> 0	0.00% (0/193)	0.56% (2/354)	1.09% (18/1647)
lbLDL-C (mg/dL)	/	/	/	/
sdLDL-C (mg/dL)	/	/	/	/
sdLDL-C/lbLDL-C	/	/	/	/

CAD: coronary artery disease; TC: total cholesterol; TG: triglyceride; HDL-C: high density lipoprotein cholesterol; LDL-C: low density lipoprotein cholesterol; sdLDL-C: small dense LDL-C; lbLDL-C: large buoyant LDL-C.

AUC values above 0.91 (Figure 5(c)), and the accuracy, precision, recall, and F1 score were above 0.87, 0.94, 0.87, and 0.92, respectively (Table 7). Interestingly, XGBoost still obtained the highest AUC (0.98), accuracy (0.95), recall (0.98), and F1 score (0.97). Moreover, KNN and XGBoost obtained the same highest F1 score (0.97) and recall (0.98). LR, SVM, and DT models all obtained the same highest precision (0.99). The results indicated that machine learning tools combing clinical features and blood lipid profile showed excellent performance to predict the CAD risk.

4. Discussion

In the past decades, a large number of studies have already reported many possible CAD risk factors, such as BMI [37], HDL-C, LDL-C, TG and TC [38], smoking, diabetes, and hypertension [39], in order to early assess the risk of CAD. Thus, in this study, we recruited 193 non-CAD controls and 2001 CAD patients and collected related risk clinical data to find out the potential risk factors for CAD. It has been reported that obesity is a common cause of cardiovascular deaths in the developed countries [40]. Moreover, diabetes has been observed to be associated with hyperlipidemia, which is characterized by increased levels of TC and decreased levels of HDL-C [41]. It has been observed that diabetic patients have higher risk of suffering from CAD than nondiabetic people [42]. Besides that, hypertension has also been frequently associated with metabolic disorders like insulin resistance or dyslipidemia, which are also known to be the risk factors of CAD [43]. Similar to the above-mentioned studies, in this study, we also found that a significantly higher prevalence of BMI, hypertension, and diabetes mellitus was observed in the CAD patients.

Besides that, increasing evidence suggests that inflammation plays an important role in the pathogenesis of CAD [44, 45]. Transforming growth factor- $\beta 1$ (TGF- $\beta 1$) is a multi-functional cytokine that regulates cell growth, differentiation, and matrix production and has a pivotal role in wound healing [46]. The high expression of TGF- $\beta 1$ level was observed during the development of many human diseases, such as periodontal disease [47] and CAD [48]. Matarese et al. reported that both TGF- $\beta 1$ and vascular endothelial growth factor (VEGF) played an important regulating role in the orchestration of the immune response in periodontal disease [47]. Although the serum level of total TGF- $\beta 1$ was upregulated in the CAD patients than in the control samples, Wang and Zhang have shown that the AUC of serum levels of total TGF- $\beta 1$ in the diagnosis of CAD was only 0.5109 [49]. Moreover, expression of VEGF is upregulated by hypoxia, inflammation, wound-healing, and other pathological processes [50]. A previous study showed that circulating levels of total VEGF-A and VEGF-A165b in CAD patients were associated with syntax score, indicating the severity and complexity of CAD [51]. Moreover, transglutaminase 2 (TG2), a protein cross-linking enzyme according to Matarese and Curro, has showed a positive correlation between TG2 and RANKL/OPG mRNA ratio, suggesting that TG2 may be involved in molecular mechanisms of inflammatory response occurring in periodontal disease [52]. In the acute myocardial infarction (AMI), model study showed that TGF- $\beta 1$ -induced transition of cardiofibroblasts into myofibroblast-like cells can be attenuated by the TG2 inhibitor 1-155, suggesting a new role for TG2 in regulating TGF- $\beta 1$ signaling in addition to its role in latent TGF- $\beta 1$ activation [53]. Besides that, cholesterol-induced sterile inflammation is thought to be central to this process via activation of a protein complex

TABLE 3: The abnormal rate of sdLDL-C subfractions in the CAD patients and non-CAD controls who were classified into LDL-C normal and high subgroups.

sdLDL-C subfractions	Control group(<i>n</i> = 193)		Non-lipid-lowering therapy group (<i>n</i> = 354)		Lipid-lowering therapy group (<i>n</i> = 1647)	
	LDL-C normal subgroup (<i>n</i> = 184)	LDL-C high subgroup (<i>n</i> = 9)	LDL-C normal subgroup (<i>n</i> = 300)	LDL-C high subgroup (<i>n</i> = 54)	LDL-C normal subgroup (<i>n</i> = 1445)	LDL-C high subgroup (<i>n</i> = 202)
	Abnormal rate	Abnormal rate	Abnormal rate	Abnormal rate	Abnormal rate	Abnormal rate
LDLC-3 (>6 mmol/L)	3.26% (6/184)	0%	78.33% (235/300)	92.59% (50/54)	82.98% (1199/1445)	95.54% (193/202)
LDLC-4 (>0 mmol/L)	1.09% (2/184)	0%	75.67% (227/300)	77.78% (42/54)	84.43% (1220/1445)	84.65% (171/202)
LDLC-5 (>0 mmol/L)	0.54% (1/184)	0%	23.33% (70/300)	29.63% (16/54)	31.35% (453/1445)	42.08% (85/202)
LDLC-6 (>0 mmol/L)	0% (0/184)	0%	4.00% (12/300)	5.56% (3/54)	5.33% (77/1445)	5.54% (11/202)
LDLC-7 (>0 mmol/L)	0% (0/184)	0%	0.67% (2/300)	0% (0/54)	1.04% (15/1445)	1.49% (3/202)

CAD: coronary artery disease; sdLDL-C: small dense LDL-C. LDL-C, low density lipoprotein cholesterol.

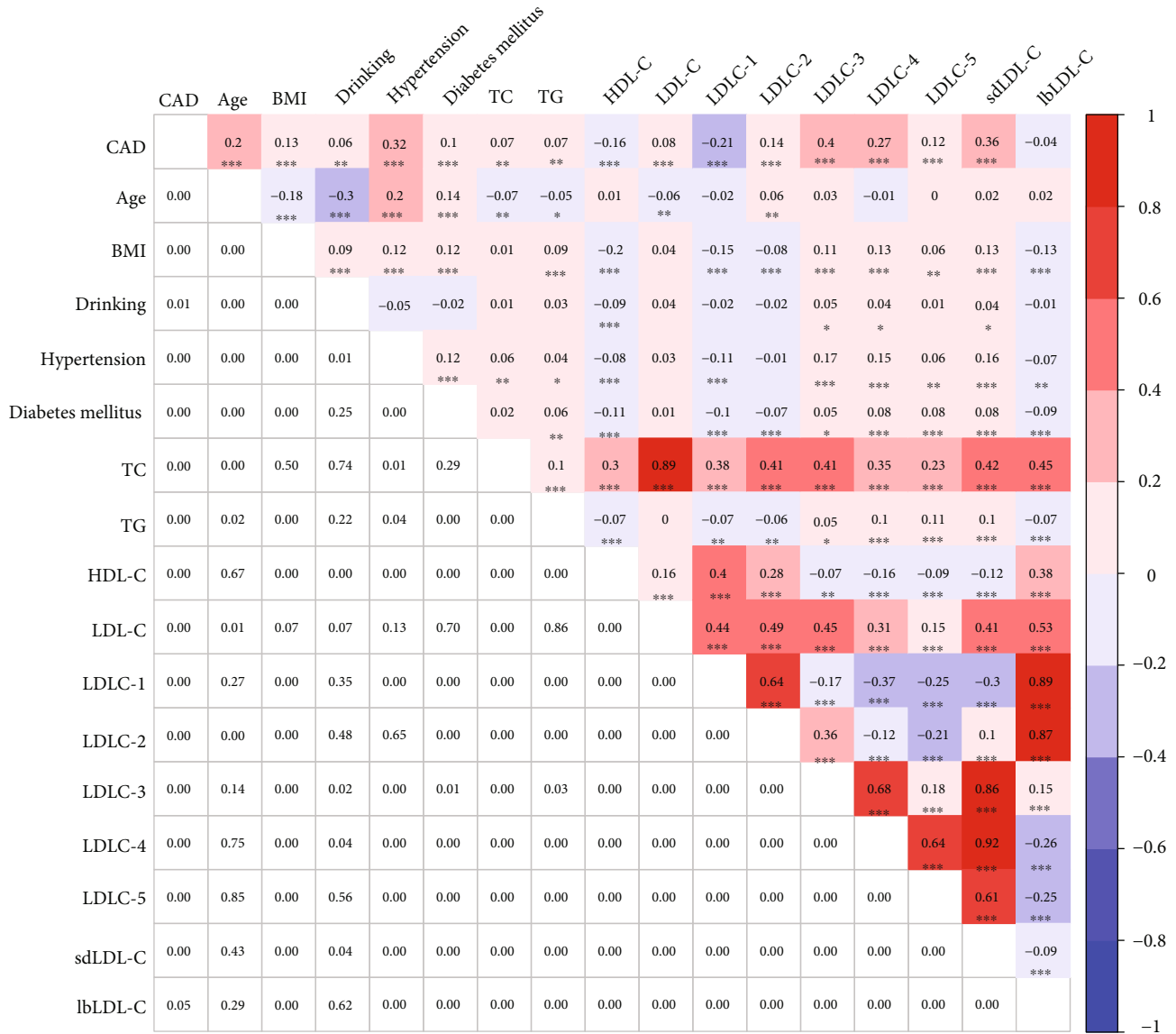
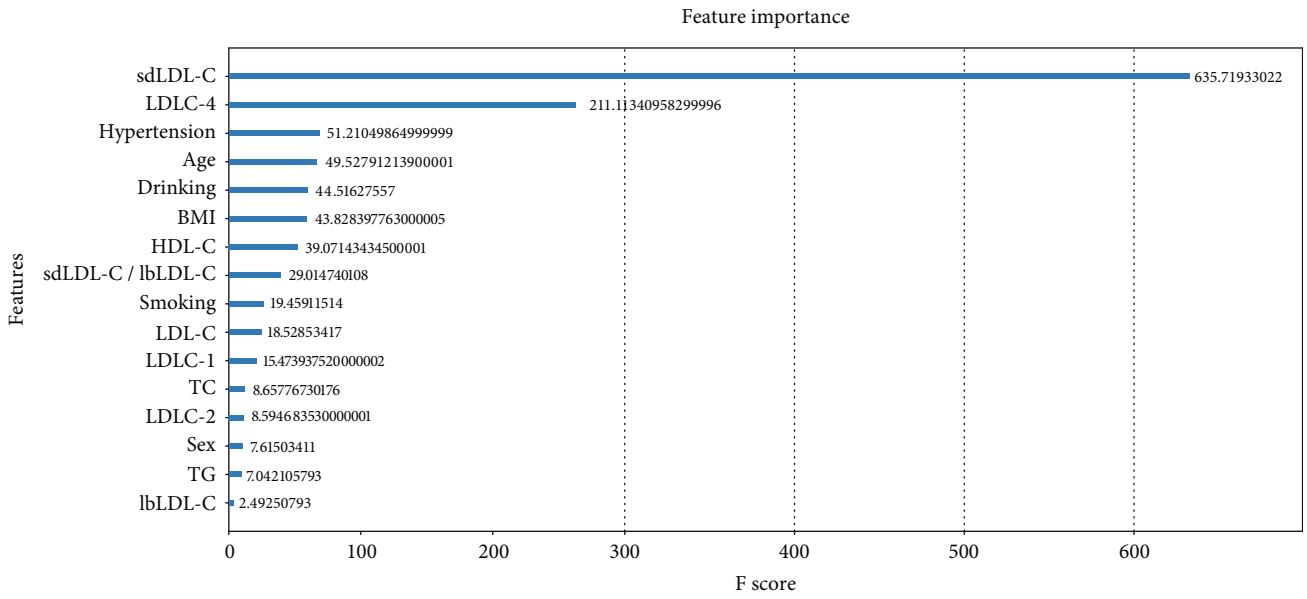


FIGURE 3: Correlation analysis among clinical characteristics, blood lipid profile, and CAD. The value in grids is Pearson correlation coefficient (Pearson’s r), which is also marked by colors. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

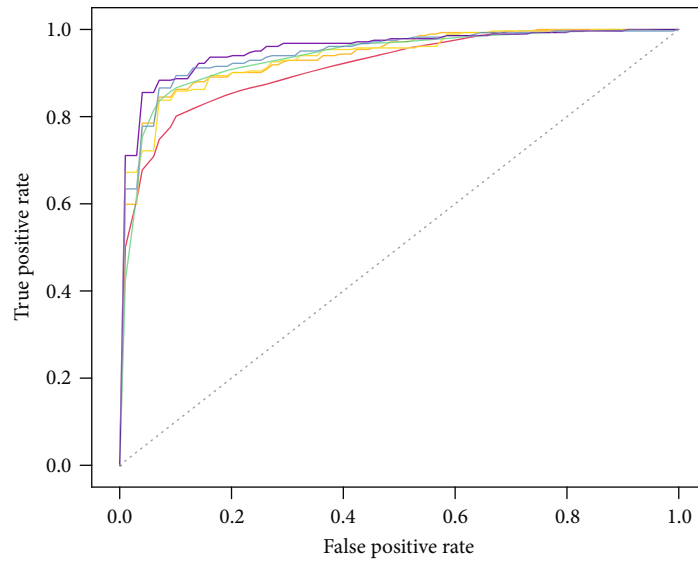
called the nucleotide-binding oligomerization domain-, leucine-rich repeat-, and pyrin domain-containing 3 (NLRP3) inflammasome. The comorbidity of smoking, hypertension, diabetes, elevated LDL-C and lipoprotein(a), or decreased HDL-C also correlated with increased NLRP3 protein expression in the aorta [54]. Zheng and Xing found that coronary atherosclerosis patients expressed high levels of NLRP3 in the aorta, which was correlated to heart disease severity [55]. However, no study has investigated whether combining these above-mentioned inflammatory factors could predict CAD risk. Hence, further researches are required to establish prediction models combining inflammatory factors to predict CAD risk in the near future. Interestingly, in recent years, the gut microbiota has been shown the capacity to contribute to substantial variation in blood lipid composition and cause CAD development [56], which can be detected by metagenomics and 16s DNA sequencing

approaches [57]. Correlations have been shown between CAD and the gut microbiota; however, the potential causal relationships are much more complex and challenging to demonstrate in the near future. In addition, we detected the traditional blood lipids in all participants and found that LDL-C level was significantly higher in the CAD patients. Large numbers of previous studies have demonstrated that LDL-C plays a crucial role in the pathogenesis and the development of CVD [58, 59]. However, CAD occurred in many patients with low LDL-C levels, and cardiovascular events even still appeared in CAD patients with intensive lipid-lowering therapy [9, 60]. Therefore, we further detected the concentrations of LDL-C subfractions by Lipoprint LDL System. Surprisingly, we found that the levels of LDLC-3 and LDL-4 were significantly higher in the CAD patients than that in the non-CAD controls. Moreover, the very high abnormal rates of LDLC-3 and LDLC-4 were found in the



(a)

ROC curves



- KNN Mean ROC (AUC = 0.91, 95%CI [0.87, 0.95])
- LR Mean ROC (AUC = 0.94, 95%CI [0.90, 0.98])
- SVM Mean ROC (AUC = 0.94, 95%CI [0.91, 0.97])
- DT Mean ROC (AUC = 0.94, 95%CI [0.91, 0.97])
- MLP Mean ROC (AUC = 0.94, 95%CI [0.92, 0.98])
- XGBoost Mean ROC (AUC = 0.95, 95%CI [0.93, 0.98])
- Chance

(b)

FIGURE 4: Continued.

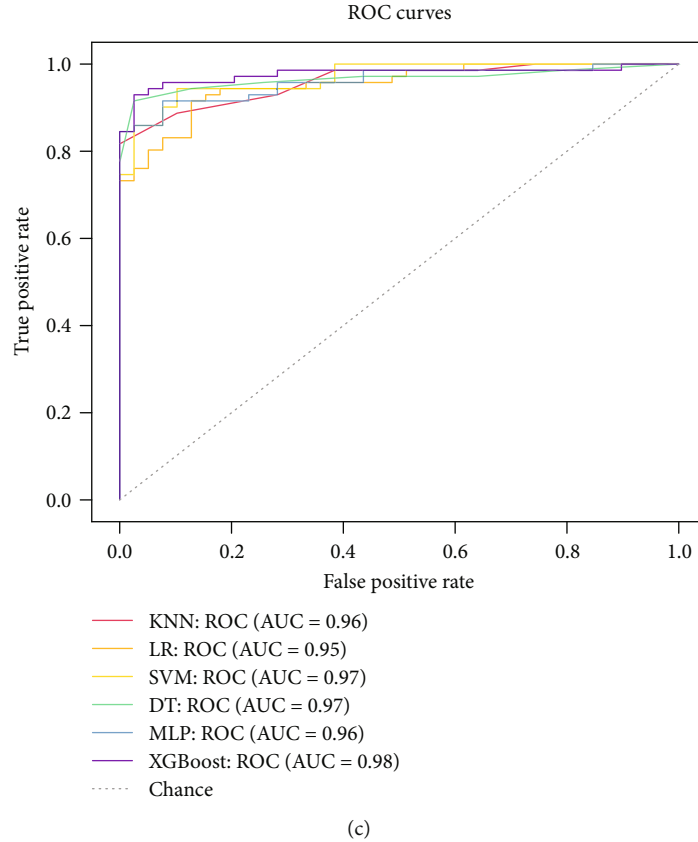


FIGURE 4: Six models were established to predict the development of risk CAD patients who did not receive any lipid-lowering treatment. The figure shows the average ROC curves of the 6 models in the training set and test set. (a) Analysis of the importance of each feature including clinical data and lipid profile in the XGBoost model. The relative importance is quantified by assigning a weight between 0 and 600 for each variable. (b) Mean AUC values and 95% CIs of all models are shown in the training set. (c) The AUC values of all models are shown in the test set. ROC: receiver operating characteristic; AUC: area under the ROC curve; CI: confidence interval; KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting.

TABLE 4: The performance of six models in the prediction of 354 CAD patients without lipid-lowering therapy in the training set.

Metrics	KNN (95% CI)	LR (95% CI)	SVM (95% CI)	DT (95% CI)	MLP (95% CI)	XGBoost (95% CI)
Accuracy	0.84 (0.79-0.88)	0.86 (0.82-0.93)	0.86 (0.83-0.89)	0.86 (0.82-0.88)	0.89 (0.85-0.93)	0.90 (0.87-0.94)
Precision	0.88 (0.84-0.91)	0.90 (0.85-0.95)	0.92 (0.90-0.94)	0.92 (0.83-0.99)	0.93 (0.91-0.94)	0.94 (0.92-0.98)
Recall	0.87 (0.80-0.91)	0.90 (0.86-0.94)	0.86 (0.823-0.88)	0.87 (0.79-0.93)	0.91 (0.86-0.96)	0.90 (0.86-0.96)
F1 score	0.87 (0.83-0.91)	0.90 (0.86-0.95)	0.89 (0.86-0.91)	0.89 (0.87-0.91)	0.92 (0.88-0.95)	0.92 (0.89-0.95)

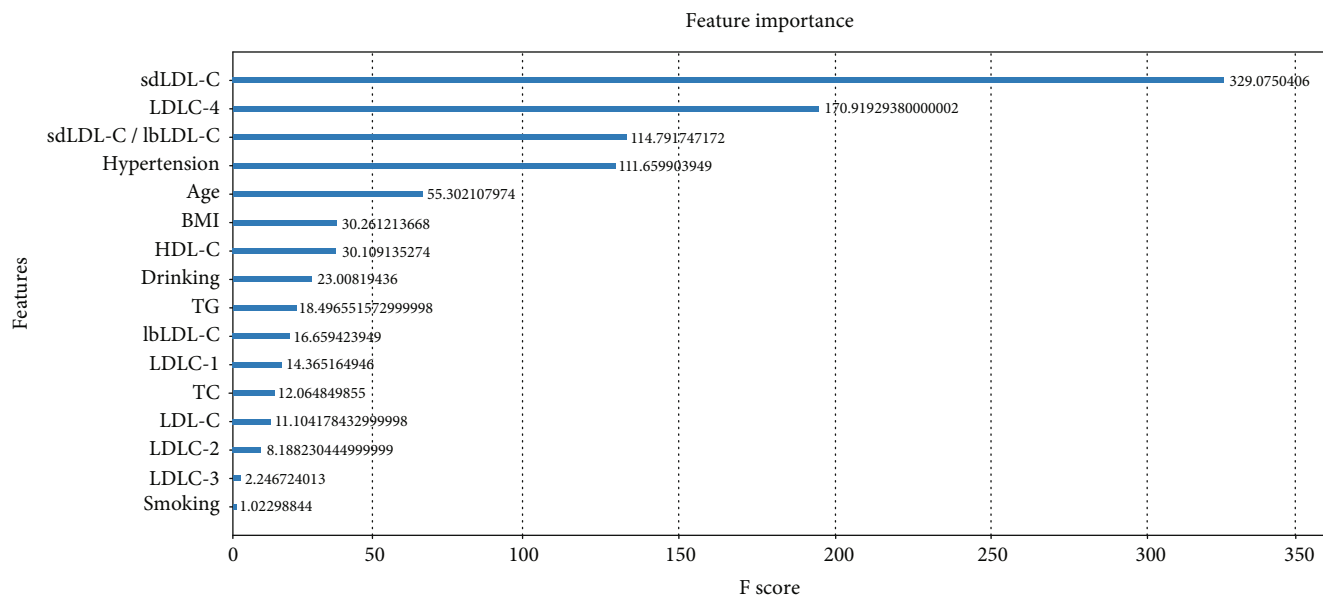
KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting; 95% CI: 95% confidence interval.

TABLE 5: The performance of six models in the prediction of 354 CAD patients without lipid-lowering therapy in the test set.

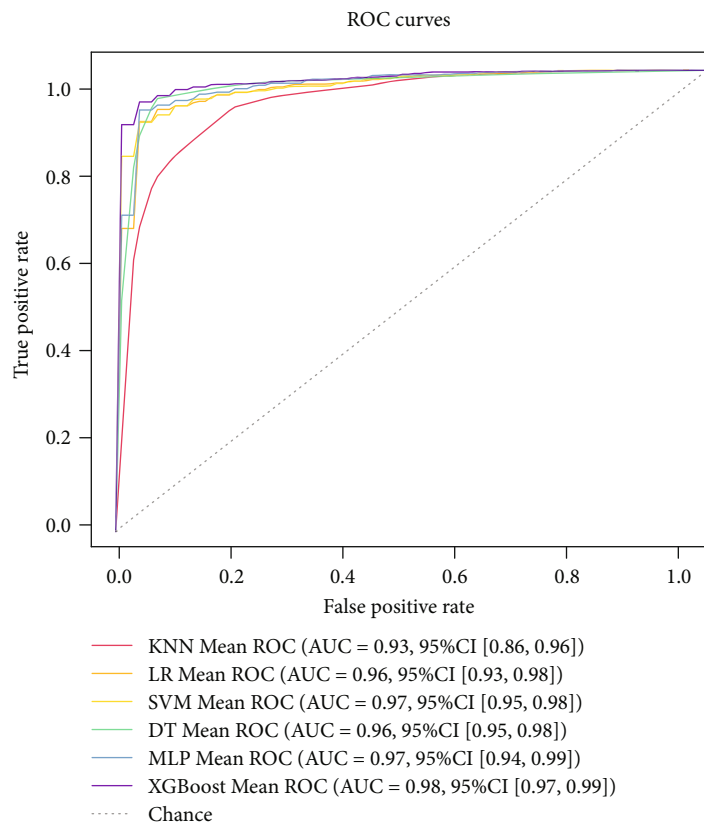
Metrics	KNN	LR	SVM	DT	MLP	XGBoost
Accuracy	0.85	0.89	0.91	0.92	0.86	0.93
Precision	0.86	0.89	0.92	0.93	0.87	0.93
Recall	0.93	0.94	0.94	0.94	0.93	0.96
F1 score	0.89	0.92	0.93	0.94	0.9	0.94

KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting.

CAD patients, even in the normal LDL-C subgroup of CAD patients. However, they are very low in the non-CAD controls. These results indicated that specific LDL-C subfractions showed powerful potential to screen high-risk CAD patients whose LDL-C levels were in normal range. Meanwhile, it can explain why many studies have reported that people with abnormal levels of sdLDL-C have high risk of cardiovascular and cerebrovascular events, even if their LDL-C levels are in normal range [61, 62]. Besides that, we found that sdLDL-C was significantly positively correlated with CAD, which is consistent with many prospective observational studies that reported sdLDL-C level was positive



(a)



(b)

FIGURE 5: Continued.

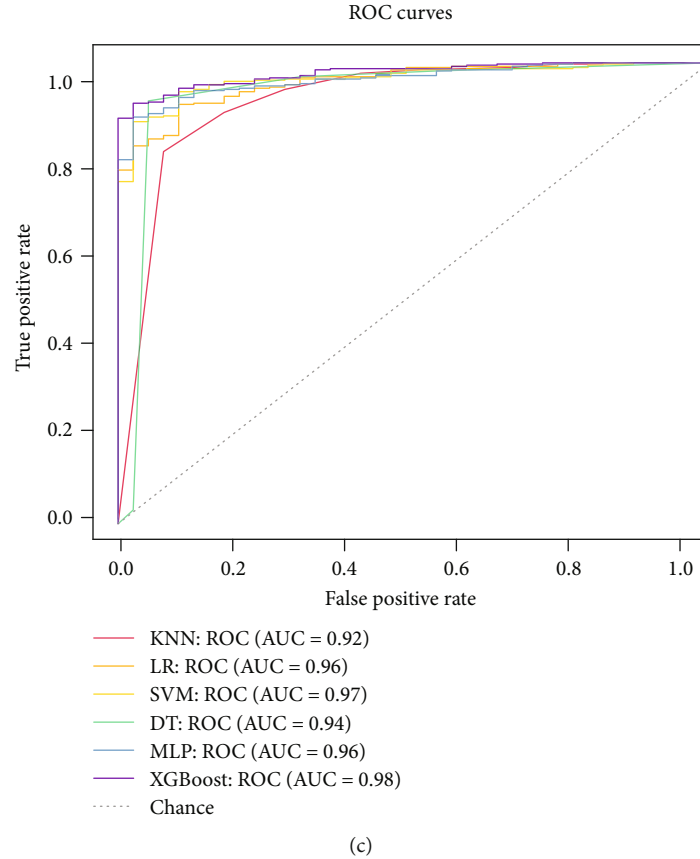


FIGURE 5: Six models were established to predict the development risk of all CAD patients. The figure shows the average ROC curves of the 6 models in the training set and test set. (a) Analysis of the importance of each feature including clinical data and lipid profile in the XGBoost model. The relative importance is quantified by assigning a weight between 0 and 350 for each variable. (b) Mean AUC values and 95% CIs of all models are shown in the training set. (c) The AUC values of all models are shown in the test set. ROC: receiver operating characteristic; AUC: area under the ROC curve; CI: confidence interval; KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting.

TABLE 6: The performance of six models in the prediction of all CAD patients in the training set.

Metrics	KNN (95% CI)	LR (95% CI)	SVM (95% CI)	DT (95% CI)	MLP (95% CI)	XGBoost (95% CI)
Accuracy	0.94 (0.93-0.95)	0.90 (0.86-0.93)	0.91 (0.89-0.93)	0.92 (0.91-0.94)	0.95 (0.94-0.96)	0.95 (0.93-0.96)
Precision	0.95 (0.94-0.95)	0.99 (0.99-1.00)	0.99 (0.99-1.00)	0.99 (0.99-1.00)	0.97 (0.96-0.98)	0.97 (0.96-0.98)
Recall	0.99 (0.98-1.00)	0.90 (0.86-0.93)	0.91 (0.90-0.93)	0.92 (0.90-0.93)	0.97 (0.96-0.98)	0.98 (0.97-0.99)
F1 score	0.97 (0.96-0.97)	0.94 (0.92-0.96)	0.95 (0.94-0.96)	0.96 (0.95-0.96)	0.97 (0.96-0.98)	0.97 (0.97-0.98)

KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting; 95% CI: 95% confidence interval.

TABLE 7: The performance of six models in the prediction of all CAD patients in the test set.

Metrics	KNN	LR	SVM	DT	MLP	XGBoost
Accuracy	0.94	0.88	0.90	0.92	0.94	0.95
Precision	0.95	0.99	0.99	0.99	0.97	0.97
Recall	0.98	0.88	0.9	0.91	0.96	0.98
F1 score	0.97	0.93	0.94	0.95	0.96	0.97

KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machine; DT: decision tree; MLP: multilayer perceptron; XGBoost: extreme gradient boosting.

association with CAD [15, 23, 63]. Meanwhile, we found that LDL-C and HDL-C were noticeably correlated with CAD too. The results suggest that LDL-C and HDL-C make the great contribution for sLDL-C composition and might be the main cause of CAD risk. However, further research studies are required to elucidate the mechanisms in the near future.

With the development of artificial intelligence (AI), machine learning is a branch of AI that describes computer models learning how to do tasks on the basis of source data rather than being rigidly programmed to do them [64, 65]. It has been attracted substantial attention for its applications in

disease diagnosis, prognosis, and treatment [34]. Machine learning has been increasingly applied in the cardiovascular research field. For example, Al'Aref and Maliakal have shown that using XGBoost model combined with coronary artery calcium scoring, age, sex, symptoms, and cardiovascular risk factors can predict obstructive CAD and yield a good AUC of 0.881 [30]. Gupta and Slater have conducted a research study that used LR, SVM, artificial neural network, and Bayesian network models combining 59 variables from real-world observational data set of 303 Iranian patients at risk for CAD. The results indicated that four models showed all AUC above 0.90 for predicting CAD [66]. However, in the above-mentioned machine learning studies for predicting CAD risk, they did not enroll sdLDL-C subfractions features. In the present study, we use six machine learning tools including KNN, LR, DT, SVM, MLP, and XGBoost-combined clinical features and blood lipid profile including sdLDL-C subfractions to predict CAD risk. We found that all models performed well in the prediction of CAD risk, which is consistent with a previous study that using above-mentioned models combining clinical data and sdLDL-C subfractions showed good performance for predicting CAD risk [3]. In addition, SVM, KNN, LR, and XGBoost models have also been reported to predict chronic kidney disease [67] and chronic obstructive pulmonary disease in Chinese population [33]. Importantly, we found that sdLDL-C ranked top 1 and LDLC-4 ranked top 2 in the feature rankings, while the study by Wu and Yang [3] indicated that sdLDL-4 ranked top 1 in the feature rankings, but sdLDL-C seems less important in that model. Therefore, large samples are needed to verify this issue in the near future. Surprisingly, although the very high abnormal rate of LDLC-3 was found in the CAD patients, it is less important in the models in our study. Interestingly, hypertension ranked top 5 in all features but ranked top 1 in the 7 clinical data. Hypertension is a well-known independent risk factor for CAD, and patients with hypertension often accompanied with abnormal lipid metabolism, which could significantly increase the risk of cardiovascular events [68, 69], which may explain why hypertension ranked higher in the model. LDL-C level cannot reflect the risk of CAD and that may explain why LDL-C ranked lower in the rankings. Meanwhile, the results indicated that sex, smoking, TC, TG, lLDL-C as well as LDLC-1, and LDLC-2 were less important in the model.

There were some limitations in this study. Firstly, the non-CAD sample size used was relatively small, and the total CAD patients and non-CAD controls was unbalanced. Secondly, we did not assess the inflammatory factors associated with CAD and input the prediction models to predict the CAD risk. Thus, further studies need to be established prediction combining-related inflammatory factors to predict CAD risk. Thirdly, these findings should be validated in a larger cohort in multicenter before these models can be applied in the clinic for CAD prediction.

In conclusion, we demonstrated that LDLC-3 and LDLC-4 were the main components of sdLDL-C and may be the main risk for CAD development. Moreover, LDLC-3, LDLC-4, and sdLDL-C were significantly positively correlated with CAD. Importantly, we identified that both

sdLDL-C and LDLC-4 play important roles in the prediction models rather than LDL-C. In addition, this study also revealed that six machine learning tools combined with clinical features and lipid profile showed excellent overall predictive power and could potentially be beneficial for the early prediction of the risk of CAD in the Chinese population.

Data Availability

The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding authors.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

TZ and SH designed the experiments and provided the resources. PX, PH, and FD performed the experiments. TZ, HT, XL, JZ, and YP acquired and analyzed the data. TZ, SH, YX, JZ, and HT wrote the manuscript. TZ, JZ, and HT revised the manuscript. All authors read and approved the final manuscript. Tiexu Zhang and Shengming Huang contributed equally to this work.

Acknowledgments

This work was supported by Shanghai's Industrial transformation and upgrading development project for absorption and innovation of imported technology (18XI-15), the Innovation Action Plan (18DZ2294600), the Shanghai Rising-Star Program (21QB1400900), and the National Natural Science Foundation of China (3210040426).

Supplementary Materials

Table S1: tuning parameters in the six predictive models. (*Supplementary Materials*)

References

- [1] GBD 2013 Mortality and Causes of Death Collaborators, "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 385, no. 9963, pp. 117-171, 2015.
- [2] M. M. Álvarez-Álvarez, D. Zanetti, R. Carreras-Torres, P. Moral, and G. Athanasiadis, "A survey of sub-Saharan gene flow into the Mediterranean at risk loci for coronary artery disease," *European Journal of Human Genetics*, vol. 25, no. 4, pp. 472-476, 2017.
- [3] D. Wu, Q. Yang, B. Su et al., "Low-density lipoprotein cholesterol 4: the notable risk factor of coronary artery disease development," *Front Cardiovascular Medicine*, vol. 8, article 619386, 2021.
- [4] Q. Wang, "Molecular genetics of coronary artery disease," *Current Opinion in Cardiology*, vol. 20, no. 3, pp. 182-188, 2005.

- [5] W. B. Kannel, "Factors of risk in the development of coronary heart disease—six year follow-up experience—Factors of risk in the development of coronary heart disease—six-year follow-up experience," *Annals of Internal Medicine*, vol. 55, no. 1, pp. 33–50, 1961.
- [6] I. Graham, D. Atar, and K. Borch-Johnsen, "European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and other societies on Cardiovascular Disease Prevention in Clinical Practice (Constituted by representatives of nine societies and by invited experts)," *European Heart Journal*, vol. 28, no. 19, pp. 2375–2414, 2007.
- [7] A. S. Go, D. Mozaffarian, V. L. Roger et al., "Heart disease and stroke statistics- 2014," *Circulation*, vol. 129, no. 3, pp. e28–e292, 2014.
- [8] P. W. F. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of coronary heart disease using risk factor Categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.
- [9] Cholesterol Treatment Trialists and E. Baigent, "Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials," *The Lancet*, vol. 376, no. 9753, pp. 1670–1681, 2010.
- [10] D. M. Hoefner, S. D. Hodel, J. F. O'Brien et al., "Development of a rapid, quantitative method for LDL subfractionation with use of the Quantimetrix Lipoprint LDL System," *Clinical Chemistry*, vol. 47, no. 2, pp. 266–274, 2001.
- [11] "Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian simvastatin survival study (4S)," *The Lancet*, vol. 344, no. 8934, pp. 1383–1389, 1994.
- [12] D. Lloyd-Jones, R. Adams, M. Carnethon et al., "Heart disease and stroke statistics—2009 update," *Circulation*, vol. 119, no. 3, p. 181, 2009.
- [13] S. Mora, N. K. Wenger, D. A. DeMicco et al., "Determinants of residual risk in secondary prevention patients treated with high- versus low-dose statin therapy," *Circulation*, vol. 125, no. 16, pp. 1979–1987, 2012.
- [14] R. C. Hoogeveen, J. W. Gaubatz, W. Sun et al., "Small dense low-density lipoprotein-cholesterol concentrations predict risk for coronary heart disease," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 34, no. 5, pp. 1069–1077, 2014.
- [15] H. Arai, Y. Kokubo, M. Watanabe et al., "Small dense low-density lipoproteins cholesterol can predict incident cardiovascular disease in an urban Japanese cohort: the Suita study," *Journal of Atherosclerosis and Thrombosis*, vol. 20, no. 2, pp. 195–203, 2013.
- [16] T. Björnheden, A. Babyi, G. Bondjers, and O. Wiklund, "Accumulation of lipoprotein fractions and subfractions in the arterial wall, determined in an in vitro perfusion system," *Atherosclerosis*, vol. 123, no. 1–2, pp. 43–56, 1996.
- [17] H. Campos, K. S. Arnold, M. E. Balestra, T. L. Innerarity, and R. M. Krauss, "Differences in receptor binding of LDL subfractions," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 16, no. 6, pp. 794–801, 1996.
- [18] B. D. Miller, E. L. Alderman, W. L. Haskell, J. M. Fair, and R. M. Krauss, "Predominance of dense low-density lipoprotein particles predicts angiographic benefit of therapy in the Stanford coronary risk intervention project," *Circulation*, vol. 94, no. 9, pp. 2146–2153, 1996.
- [19] P. Srisawasdi, S. Vanavanan, M. Rochanawutanon, K. Kruthkul, K. Kotani, and M. H. Kroll, "Small-dense LDL/large-buoyant LDL ratio associates with the metabolic syndrome," *Clinical Biochemistry*, vol. 48, no. 7–8, pp. 495–502, 2015.
- [20] Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, "Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)," *JAMA: The Journal of the American Medical Association*, vol. 285, no. 19, pp. 2486–2497, 2001.
- [21] A. L. Catapano, I. Graham, G. De Backer et al., "2016 ESC/EAS guidelines for the management of dyslipidaemias," *European Heart Journal*, vol. 37, no. 39, pp. 2999–3058, 2016.
- [22] M. Y. Tsai, B. T. Steffen, W. Guan et al., "New automated assay of small dense low-density lipoprotein cholesterol identifies risk of coronary heart disease," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 34, no. 1, pp. 196–201, 2014.
- [23] M. Higashioka, S. Sakata, T. Honda et al., "Small dense low-density lipoprotein cholesterol and the risk of coronary heart disease in a Japanese community," *Journal of Atherosclerosis and Thrombosis*, vol. 27, no. 7, pp. 669–682, 2020.
- [24] M. Balling, B. G. Nordestgaard, A. Langsted, A. Varbo, P. R. Kamstrup, and S. Afzal, "Small dense low-density lipoprotein cholesterol predicts atherosclerotic cardiovascular disease in the Copenhagen general population study," *Journal of the American College of Cardiology*, vol. 75, no. 22, pp. 2873–2875, 2020.
- [25] R. B. D'Agostino, R. S. Vasan, M. J. Pencina et al., "General cardiovascular risk profile for use in primary care," *Circulation*, vol. 117, no. 6, pp. 743–753, 2008.
- [26] S. M. Grundy, N. J. Stone, A. L. Bailey et al., "2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APHA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines," *Journal of the American College of Cardiology*, vol. 73, no. 24, pp. 3168–3209, 2019.
- [27] P. M. Ridker, E. Danielson, F. A. H. Fonseca et al., "Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein," *The New England Journal of Medicine*, vol. 359, no. 21, pp. 2195–2207, 2008.
- [28] K. Nasir, E. D. Michos, R. S. Blumenthal, and P. Raggi, "Detection of high-risk young adults and women by coronary calcium and national cholesterol education program panel III guidelines," *Journal of the American College of Cardiology*, vol. 46, no. 10, pp. 1931–1936, 2005.
- [29] B. Yang, S. Yan, J. Yan et al., "A study of the association of rs12040273 with susceptibility and severity of coronary artery disease in a Chinese Han population," *BMC Cardiovascular Disorders*, vol. 18, no. 1, p. 10, 2018.
- [30] S. J. Al'Aref, G. Maliakal, G. Singh 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," *European Heart Journal*, vol. 41, no. 3, pp. 359–367, 2020.
- [31] I. Rajman, M. J. Kendall, R. Cramb, R. L. Holder, M. Salih, and M. D. Gammage, "Investigation of low density lipoprotein subfractions as a coronary risk factor in normotriglyceridaemic men," *Atherosclerosis*, vol. 125, no. 2, pp. 231–242, 1996.

- [32] China joint committee on revision of guidelines for prevention and treatment, "Chinese guideline for the management of dyslipidemia in adults," *Chinese Circulation Journal*, vol. 31, no. 10, pp. 937–950, 2016.
- [33] X. Ma, Y. Wu, L. Zhang et al., "Comparison and development of machine learning tools for the prediction of chronic obstructive pulmonary disease in the Chinese population," *Journal of Translational Medicine*, vol. 18, no. 1, p. 146, 2020.
- [34] N. Radakovich, M. Nagy, and A. Nazha, "Machine learning in haematological malignancies," *The Lancet Haematology*, vol. 7, no. 7, pp. e541–e550, 2020.
- [35] Z. Wang, M. Hu, and G. Zhai, "Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data," *Sensors*, vol. 18, no. 4, p. 1126, 2018.
- [36] A. H. Taylor, M. Finney, and P. M. Lam, "Modulation of the endocannabinoid system in viable and non-viable first trimester pregnancies by pregnancy-related hormones," *Reproductive Biology and Endocrinology*, vol. 9, no. 1, p. 152, 2011.
- [37] A. R. Wood, T. Esko, J. Yang et al., "Defining the role of common variation in the genomic and biological architecture of adult human height," *Nature Genetics*, vol. 46, no. 11, pp. 1173–1186, 2014.
- [38] C. J. Willer, E. M. Schmidt, and S. Sengupta, "Discovery and refinement of loci associated with lipid levels," *Nature Genetics*, vol. 45, no. 11, pp. 1274–1283, 2013.
- [39] A. K. Malakar, D. Choudhury, and B. Hadler, "A review on coronary artery disease, its risk factors, and therapeutics," *Journal of Cellular Physiology*, vol. 234, no. 10, pp. 16812–16823, 2019.
- [40] Y. Matsuzawa, T. Nakamura, I. Shimomura, and K. Kotani, "Visceral fat accumulation and cardiovascular disease," *Obesity Research*, vol. 3, no. S5, pp. 645S–647S, 1995.
- [41] S. M. Haffner, "Diabetes, hyperlipidemia, and coronary artery disease," *The American Journal of Cardiology*, vol. 83, no. 9, pp. 17–21, 1999.
- [42] L. E. Wagenknecht, R. B. D'Agostino Jr., and S. M. Haffner, "Impaired glucose tolerance, type 2 diabetes, and carotid wall thickness: the insulin resistance atherosclerosis study," *Diabetes Care*, vol. 21, no. 11, pp. 1812–1818, 1998.
- [43] R. A. DeFronzo and E. Ferrarinni, "Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease," *Diabetes Care*, vol. 14, no. 3, pp. 173–194, 1991.
- [44] K. Sugamura and J. F. Keaney, "Reactive oxygen species in cardiovascular disease," *Free Radical Biology & Medicine*, vol. 51, no. 5, pp. 978–992, 2011.
- [45] D. Tousoulis, A. M. Kampoli, and N. Papageorgiou, "Pathophysiology of atherosclerosis: the role of inflammation," *Current Pharmaceutical Design*, vol. 17, no. 37, pp. 4089–4110, 2011.
- [46] G. Matarese, G. Isola, G. P. Anastasi et al., "Immunohistochemical analysis of TGF- β 1 and VEGF in gingival and periodontal tissues: a role of these biomarkers in the pathogenesis of scleroderma and periodontal disease," *International Journal of Molecular Medicine*, vol. 30, no. 3, pp. 502–508, 2012.
- [47] G. Matarese, G. Isola, G. P. Anastasi et al., "Transforming growth factor beta1 and vascular endothelial growth factor levels in the pathogenesis of periodontal disease," *European Journal of Inflammation*, vol. 11, no. 2, pp. 479–488, 2013.
- [48] C. Chen, W. Lei, W. Chen et al., "Serum TGF- β 1 and SMAD3 levels are closely associated with coronary artery disease," *BMC Cardiovascular Disorders*, vol. 14, no. 1, p. 18, 2014.
- [49] S. Wang, Q. Zhang, Y. Wang et al., "Transforming growth factor β 1 (TGF- β 1) appears to promote coronary artery disease by upregulating sphingosine kinase 1 (SPHK1) and further upregulating its downstream TIMP-1," *Medical Science Monitor*, vol. 24, pp. 7322–7328, 2018.
- [50] S. Ramakrishnan, V. Anand, and S. Roy, "Vascular endothelial growth factor signaling in hypoxia and inflammation," *Journal of neuroimmune pharmacology*, vol. 9, no. 2, pp. 142–160, 2014.
- [51] Y. Shibata, "Balance between angiogenic and anti-angiogenic isoforms of VEGF-A is associated with the complexity and severity of coronary artery disease," *Clinica chimica acta*, vol. 478, pp. 114–119, 2018.
- [52] G. Matarese, M. Currò, G. Isola et al., "Transglutaminase 2 up-regulation is associated with RANKL/OPG pathway in cultured HPDL cells and THP-1-differentiated macrophages," *Amino Acids*, vol. 47, no. 11, pp. 2447–2455, 2015.
- [53] Z. Wang, D. J. Stuckey, C. E. Murdoch, P. Camelliti, G. Y. H. Lip, and M. Griffin, "Cardiac fibrosis can be attenuated by blocking the activity of transglutaminase 2 using a selective small-molecule inhibitor," *Cell Death & Disease*, vol. 9, no. 6, p. 613, 2018.
- [54] G. Hemenway and W. H. Frishman, "Therapeutic implications of NLRP3-mediated inflammation in coronary artery disease," *Cardiology in Review*, vol. 30, no. 2, pp. 90–99, 2022.
- [55] F. Zheng, S. Xing, Z. Gong, and Q. Xing, "NLRP3 inflammasomes show high expression in aorta of patients with atherosclerosis," *Heart, Lung & Circulation*, vol. 22, no. 9, pp. 746–750, 2013.
- [56] J. Fu, M. J. Bonder, M. C. Cenit et al., "The gut microbiome contributes to a substantial proportion of the variation in blood lipids," *Circulation Research*, vol. 117, no. 9, pp. 817–824, 2015.
- [57] L. Martellacci, G. Quaranta, R. Patini, G. Isola, P. Gallenzi, and L. Masucci, "A literature review of metagenomics and culturomics of the peri-implant microbiome: current evidence and future perspectives," *Materials (Basel)*, vol. 12, no. 18, p. 3010, 2019.
- [58] S. M. Grundy, N. J. Stone, A. L. Bailey 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*, vol. 139, no. 25, pp. e1082–e1143, 2019.
- [59] F. Mach, C. Baigent, A. L. Catapano et al., "2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk," *European Heart Journal*, vol. 41, no. 1, pp. 111–188, 2020.
- [60] G. Heart Protection Study Collaborative, "MRC/BHF, "Heart protection study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial," *The Lancet*, vol. 360, no. 9326, pp. 7–22, 2002.
- [61] W. J. Storkus, R. D. Salter, P. Cresswell, and J. R. Dawson, "Peptide-induced modulation of target cell sensitivity to natural killing," *Journal of Immunology*, vol. 149, no. 4, pp. 1185–1190, 1992.
- [62] E. D. Michos, J. W. McEvoy, and R. S. Blumenthal, "Lipid management for the prevention of atherosclerotic

- cardiovascular disease,” *The New England Journal of Medicine*, vol. 381, no. 16, pp. 1557–1567, 2019.
- [63] Y. Zhang, R. X. Xu, S. Li et al., “Association of plasma small dense LDL cholesterol with PCSK9 levels in patients with angiographically proven coronary artery disease,” *Nutrition, Metabolism, and Cardiovascular Diseases*, vol. 25, no. 4, pp. 426–433, 2015.
- [64] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [65] V. H. Buch, I. Ahmed, and M. Maruthappu, “Artificial intelligence in medicine: current trends and future possibilities,” *The British Journal of General Practice*, vol. 68, no. 668, pp. 143–144, 2018.
- [66] A. Gupta, J. J. Slater, D. Boyne et al., “Probabilistic graphical modeling for estimating risk of coronary artery disease: applications of a flexible machine-learning method,” *Medical Decision Making*, vol. 39, no. 8, pp. 1032–1044, 2019.
- [67] J. Xiao, R. Ding, X. Xu et al., “Comparison and development of machine learning tools in the prediction of chronic kidney disease progression,” *Journal of Translational Medicine*, vol. 17, no. 1, p. 119, 2019.
- [68] H. H. Liu, Y. X. Cao, S. Li et al., “Impacts of prediabetes mellitus alone or plus hypertension on the coronary severity and cardiovascular outcomes,” *Hypertension*, vol. 71, no. 6, pp. 1039–1046, 2018.
- [69] S. S. Franklin and N. D. Wong, “Hypertension and cardiovascular disease: contributions of the Framingham heart study,” *Heart*, vol. 8, no. 1, pp. 49–57, 2013.