BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Unlocking the link: predicting cardiovascular disease risk with a focus on airflow obstruction using machine learning

Xiyu Cao[1†], Jianli Ma[1†], Xiaoyi He[2†], Yufei Liu[1], Yang Yang[3], Yaqi Wang[1] and Chuantao Zhang[1*]

Abstract

Background Respiratory diseases and Cardiovascular Diseases (CVD) often coexist, with airflow obstruction (AO) severity closely linked to CVD incidence and mortality. As both conditions rise, early identification and intervention in risk populations are crucial. However, current CVD risk models inadequately consider AO as an independent risk factor. Therefore, developing an accurate risk prediction model can help identify and intervene early.

Methods This study used the National Health and Nutrition Examination Survey (NHANES) III (1988–1994) and NHANES 2007–2012 datasets. Inclusion criteria were participants aged over 40 with complete AO and CVD data; exclusions were those with missing key data. Analysis included 12 variables: age, gender, race, PIR, education, smoking, alcohol, BMI, hyperlipidemia, hypertension, diabetes, and AO. Logistic regression analyzed the association between AO and CVD, with sensitivity and subgroup analyses. Six ML models predicted CVD risk for the general population, using AO as a predictor. RandomizedSearchCV with 5-fold cross-validation was used for hyperparameter optimization. Models were evaluated by AUC, accuracy, precision, recall, F1 score, and Brier score, with the SHapley Additive exPlanations (SHAP) enhancing explainability. A separate ML model was built for the subpopulation with AO, evaluated similarly.

Results The cross-sectional analysis showed that there was a significant positive correlation between AO occurrence and CVD prevalence, indicating that AO is an important risk factor for CVD (all $P < 0.05$). For the general population, the XGBoost model was selected as the optimal model for predicting CVD risk (AUC = 0.7508, AP = 0.3186). The top three features in terms of importance were age, hypertension, and PIR. For the subpopulation with airflow obstruction, the XGBoost model was also selected as the optimal model for predicting CVD risk (AUC = 0.6645, AP = 0.3545). SHAP shows that education level has the greatest impact on predicting CVD risk, followed by gender and race.

Conclusion AO correlates positively with CVD. Age, hypertension, PIR affect CVD risk most in general. For AO patients, education, gender, ethnicity are key CVD risk factors.

Keywords Cardiovascular disease, Airflow obstruction, Co-morbidity, Machine learning, Prediction model

†Xiyu Cao, Jianli Ma and Xiaoyi He contributed equally to this work and share first authorship.

*Correspondence:
Chuantao Zhang
zhangchuantao@cdutcm.edu.cn

[1]Department of Respiratory Medicine, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China
[2]Columbia University, New York, NY, USA
[3]Department of Gastroenterology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China

Introduction

Cardiovascular Disease (CVD) is a group of diseases involving the heart or blood vessels including coronary heart disease (CHD), myocardial infarction and stroke [1–3]. CVD poses a major threat to human health and is the leading cause of morbidity and mortality globally [4, 5]. 17.9 million people are estimated to die from CVD in 2019, accounting for 32% of all deaths globally, 85% of which are from heart attack and stroke [6]. Over the past decade, global CVD deaths have climbed by 12.5% [7]. The overall disease burden of CVD is heavy, with increasing incidence and high prevalence and mortality rates, especially in older populations and those with chronic co-morbidities [8, 9]. Most of the CVD can be prevented and ameliorated by lifestyle changes and medical interventions, such as reducing smoking, modifying diet, controlling body weight, increasing physical activity and avoiding alcohol abuse [6, 10]. Therefore, an in-depth understanding of the risk factors for CVD is important for the prevention and control of CVD.

Comorbidity refers to the coexistence of multiple diseases in an individual, which is associated with poor quality of life, polypharmacy, and high mortality rates [11]. Respiratory diseases and CVD frequently coexist [12]. Airflow Obstruction (AO) is a common respiratory problem that is characterised by obstruction of airflow in the airways, leading to dyspnoea and reduced lung function. AO comprises variable obstruction and fixed obstruction, with the latter satisfying the definition of chronic obstructive pulmonary disease (COPD) under the GOLD criteria [13]. In recent years, a growing body of research has shown a strong association between AO and CVD. Several studies have found a positive correlation between AO severity and the risk of CVD [14–19]. Impairment of several lung function indices, such as forced expiratory volume in 1 s (FEV1) and forced vital capacity (FVC), has been reported as an independent predictor of CVD [20]. Reduced COPD, FEV1, FVC and FEV1/FVC ratio have all been reported to be associated with an increased risk of coronary artery disease [21, 22]. There is evidence that COPD and impaired lung function are associated with an increased risk of stroke [23]. Reduced lung function indices such as FEV1 and FVC are negatively associated with CHD, stroke and other CVD mortality [24, 25]. There may be multiple potential mechanisms for this association. In addition to traditional cardiovascular risk factors, inflammation is an important risk factor for CVD [26]. Impaired lung function and associated lung diseases can have a direct deleterious effect on cardiovascular health through a variety of biological pathways, including systemic inflammation or oxidative stress [27, 28]. Furthermore, the combined effects of airflow obstruction and chronic diseases may also lead to vascular injury, increasing the risk of atherosclerosis and CVD [29].

Existing studies have shown a close relationship between the severity of AO and CVD prevalence and mortality [30, 31]. However, while these studies have revealed the association between AO and CVD, they primarily focus on epidemiological evidence and do not provide a comprehensive cardiovascular disease risk prediction model [30, 31].

In recent years, with the advancement of machine learning (ML) technologies, research has focused on developing efficient and accurate CVD prediction models. Studies have improved the accuracy and efficiency of CVD diagnosis through data preprocessing using min-max scaling and hyperparameter tuning with Bayesian optimisation [32]. Previous research explored the application of four tree-based machine learning methods in CVD prediction and five variables (age, LDL, history of cardiac disease in first-degree relatives, physical activity level, and hypertension status) were found to be the most influential [33]. Age, sex, race, socioeconomic status, lifestyle, and clinical factors have all been considered significant predictors of CVD [33–37]. However, no CVD prediction model has yet AO as a predictive factor, and there is a lack of CVD prediction models specifically for populations with AO. This study fills this gap, enhancing the accuracy of risk prediction and aiding in the early identification and intervention of CVD risk in individuals with AO.

In this study, we used the National Health and Nutrition Examination Survey (NHANES III and 2007–2012) dataset to investigate the relationship between airflow obstruction and CVD. Subsequently, six machine learning (ML) models were constructed to predict CVD risk in the general population, with AO as one of the predictive factors. Additionally, we conducted a separate analysis for the AO subpopulation to predict their CVD risk. SHapley Additive exPlanations (SHAP) was used to determine the contribution of each variable to CVD identification. This study enhances the potential for early intervention and is expected to provide practical guidance for the development of prevention and management strategies for specific populations.

Method

Study design and population

This study pooled data from the NHANES III and NHANES 2007–2012. NHANES is a population-based national survey that measures the health and nutritional status of the American general population every 2 years using questionnaires, physical examination, and biospecimen collection. Detailed study procedures of NHANES have been described by the Centers for Disease Control and Prevention (CDC). All NHANES studies passed the National Center for Health Statistics (NCHS) Ethics Review Board and written informed consent was

obtained from all participants (https://www.cdc.gov/nchs/nhanes/irba98.htm).

NHANES III

NHANES III, a cross-sectional survey with a complex, stratified, multistage probability cluster sampling design, was conducted from 1988 to 1994. We obtained 20,050 valid samples from NHANES III.

NHANES 2007–2012

NHANES 2007–2012 comprises consecutive cross-sectional surveys, each spanning approximately two years. We obtained 30,442 valid samples from the three cycles.

The differences between NHANES III and NHANES 2007–2012 are significant, primarily in terms of time span, technological advancements, socio-economic changes, and data integrity. During this period, advancements in detection techniques and methodologies have influenced the measurement of health indicators. Changes in the economic condition of the United States may have impacted participants' health status and behavioral patterns.

Measurements

In this study, participants aged ≥ 40 years were included. Subjects with missing data on important variables such as FEV1, FVC data for pulmonary function, CVD-related data, age, gender, race, ratio of family income to poverty (PIR), education level, smoking status, alcohol status, BMI, blood lipid related data, hypertension data/problems, and fasting blood glucose level/self-reported diabetes history/use of oral hypoglycemic drugs/use of insulin/hemoglobin were excluded from the analysis. Ultimately, the study included 12,052 participants, of which 6,517 were from NHANES III and 5,535 were from NHANES 2007–2012 (Fig. 1).

Assessment of spirometry

Detailed information on lung capacity measurement equipment, examination plans, calibration procedures, and quality control in NHANES is available (https://wwwn.cdc.gov/nchs/nhanes/nhanes3/Default.aspx). The FEV1 and FVC values specified by each subject are determined by the maximum values of FEV1 and FVC in the lung capacity measurement of each subject, respectively. The testing program complies with the recommendations of the American Thoracic Society [38]. In this study, FEV1 and FVC values were included to define AO. AO is defined as FEV1/FVC < 0.7% [13, 39].

Assessment of CVD

CVD was defined as a combination of self-reported physician diagnosis of CHD, myocardial infarction, angina, congestive heart failure (HF), or stroke. All participants were asked the following question: ''Has a doctor or other health professional ever told you that you had congestive HF/ CHD /angina, also called angina pectoris/ a heart attack (also called myocardial infarction)/a stroke?". If any of the above questions answer yes the participant was considered to have CVD [40]. In addition, in NHANES

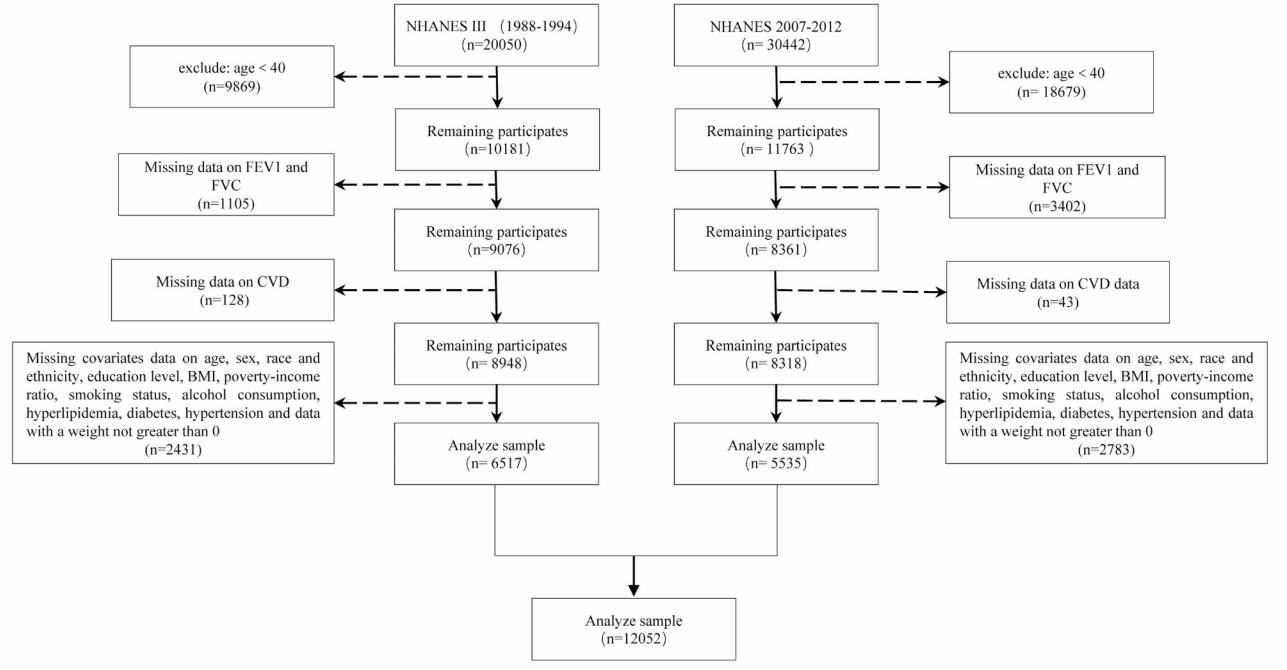

Fig. 1 Flowchart of the study population selection from NHANES III and NHANES 2007–2012

III, angina was defined as a positive result on the ROSE questionnaire, and CHD was defined as a combination of self-reported physician diagnosis of myocardial infarction and angina (Figure S1) [41].

Covariates

Based on both existing literature and clinical insights [42], this study included the following covariates: age, gender, race, PIR, education level, smoking status, alcohol status, BMI, hypertension, hyperlipidemia, and diabetes. The above profile information was collected from demographic and censored public information released by NHANES. This study categorized race into four categories: Non-Hispanic White, Non-Hispanic Black, Mexican American, and Other Race. PIR was categorized as low (≤ 1.3), medium (1.3–3.5), and high (> 3.5) based on the household poverty income ratio. Likewise, educational level was categorized as low (Less than 9th grade), middle [9-11th grade (Includes 12th grade with no diploma)], high (High school or equivalent and Some college or more). Smoking status was determined by NHANES survey questions, and participants were defined as smokers if they had smoked at least 100 cigarettes in their lifetime. Drinking status was categorised as never (< 12 drinks in a lifetime), moderate (< = 2 drinks/day for men and < = 1 drink/day for women) and over (> 2 drinks/day for men and > 1 drink/day for women). Alcohol consumption was calculated as: alcohol consumption = (average frequency of alcohol consumption per year × average daily alcohol consumption)/365 days. BMI was categorised as underweight (< 18.5), normal weight (18.5–24.9), overweight (25.0-29.9), and obesity (≥ 30.0 kg/m^2). Hyperlipidemia was defined as a serum total cholesterol of 200 mg/dL, or triglycerides of 150 mg/ dL, or HDL of 40 mg/dL for men and 50 mg/dL for women, or LDL of 130 mg/dL. Hypertension was defined as three times mean systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg, or a self-reported history of hypertension or current use of prescription medication for HBP. Diabetes was defined as a fasting blood glucose level ≥ 126 mg/dL or a HbA1c of ≥ 6.5% or use of oral hypoglycaemic drugs or use of insulin or a self-reported history of diabetes.

Logistic regression and subgroup

Firstly, multifactorial survey-weighted logistic regression was performed in this study to discuss the effects of AO and FEV1/FVC on CVD risk. Also, sensitivity analyses using quartiles of FEV1/FVC as categorical variables were performed, and multifactorial survey-weighted logistic regression was again performed. Model 1, included only AO as the independent variable without considering other confounding factors. Model 2 extended Model 1 by incorporating age, gender, race, PIR, and education level as covariates to control their influence on the AO-CVD association. Model 3, building upon Model 2, further included all possible confounding factors such as smoking status, alcohol consumption, BMI, hyperlipidemia, hypertension, and diabetes, comprehensively adjusting these covariates to accurately assess the association between AO and CVD. We also examined the variance inflation factor (VIF) of the models to detect multicollinearity.

Further, to investigate whether this relationship was altered by age, gender, race, PIR, education, BMI, smoking status, drinking status, Hyperlipidemia, hypertension and diabetes, we performed interaction and subgroup analyses of the presence of AO and FEV1/FVC, respectively, to test the stability of our results.

Machine learning models for CVD diagnosis

The general population CVD model used 12 features (age, gender, race, PIR, BMI, smoking status, drinking status, Hyperlipidemia, hypertension and diabetes), while the AO population CVD model used 11 features, excluding AO. The data for both were randomly partitioned into training and test sets in the ratio of 8:2. For general population CVD model, six ML models, LightGBM, XGBoost, Naive Bayes, KNN, Random Forest, and CatBoost, were used to predict CVD. For AO population CVD model, six ML models, LightGBM, XGBoost, Naive Bayes, KNN, Random Forest, and MLP, were used. Metrics such as Area Under the Curve (AUC), PR curve, Accuracy, Precision, Recall, Brier Score, and F1 Score were used to evaluate the ML models.

Class balancing and feature selection methods

To balance the class distribution, the CVD model adopted a replication-based oversampling technique, which entailed replicating minority class samples until their count equalled that of the majority, thus creating a balanced dataset. The AO population model used the Synthetic Minority Over-Sampling Technique (SMOTE) to identify the minority class samples and expand their number to match the majority class, selecting an appropriate number of neighbours to ensure the generated samples were representative and realistic. The method for feature selection involved two independent processes: forward selection and backward elimination. Forward selection starts by adding one feature at a time, beginning with none, while backward elimination starts with all features and removes them one by one. Each addition or removal aims for a maximum increase in performance. The results from both methods are compared afterward to determine the optimal set of features. It helps explore the feature space more thoroughly.

Model validation and hyperparameter tuning

To accurately assess the model's generalization performance and reduce random bias, we employ 5-fold cross-validation, dividing the training set into 5 subsets, using 4 for training and 1 for validation in rotation, and averaging the performance metrics over 5 iterations. For hyperparameter optimization, RandomizedSearchCV was used to hunt for the best hyperparameter combinations for all models, with 5-fold cross-validation and 100 iterations. For tree-based models such as LightGBM and XGBoost, we particularly limiting the number of leaf nodes and maximum depth to no more than 6 to prevent overfitting. This prevents the model from fitting noise in the training data, which helps avoid overfitting.

SHAP-based model interpretation

The SHAP algorithm was used to provide global and local interpretations for the best-performing models. Global interpretations can provide consistent and accurate attribute values for each feature in the model to show associations between input features and CVDs. Local interpretations can demonstrate specific predictions for individual samples by inputting specific data.

The statistical analysis part was completed by R (4.3.1). Bilateral $P<0.05$ was considered statistically different. The modeling and verification part of ML was conducted in Python (3.10.12). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.

Results

Baseline characteristics of study participants

In this study, a total of 12,052 subjects were enrolled and the screening process is shown in Fig. 1. Table 1 shows the demographic characteristics of all participants. The participants were divided into two groups according to the presence or absence of CVD, and there were significant differences in age, gender, PIR, education, BMI, smoking status, hypertension, diabetes, FEV1, FVC, FEV1/FVC, and AO ($P<0.05$). Population with CVD were older, mainly concentrated in the age group of 60–79 years, and were more likely to be male, with a lower PIR, education, and FEV1/FVC. In addition, they also had higher BMI, with the highest percentage of obesity. Compared to the no-CVD population, the percentages of smoking, hypertension, diabetes and AO were significantly higher. Table S1 shows the demographic characteristics of the AO-only population. Compared with patients without CVD, patients with combined CVD showed similar characteristics to those in Table 1.

The association of candidate predictor variables and dependent variables

To reveal the relationship between AO and CVD risk, three logistic regression models were constructed in this study by adjusting for different confounding variables to assess the association (Table 2). There was a significant positive association between AO and the prevalence of CVD in the original model 1, in the partially adjusted model 2, and in model 3 adjusted for all confounding variables (all $P<0.05$). In Model 1, individuals with AO had a 2.08 times higher risk of developing CVD compared to those without AO (OR = 2.08, 95% CI: 1.79, 2.42, $P<0.001$). After adjusting for confounding factors, a significant association between AO and CVD remained, with OR values (95% CI) of 1.38 (1.15, 1.65) in Model 2 and 1.33 (1.10, 1.61) in Model 3. Consistent with this result, when FEV1/FVC was considered as a categorical variable (quartiles) for sensitivity analysis, there was a significant association between FEV1/FVC and CVD (all $P<0.05$) (Table S2). As shown in Figure S2, the VIF were all between 1 and 2, indicating no multicollinearity between the variables.

We performed further stratified analyses to assess the impact of AO on CVD (Table 3). Most of the variables, including age, gender, race, BMI, smoking status, drinking status, Hyperlipidemia, hypertension and diabetes did not significantly alter the correlation between AO and CVD (all P for interaction >0.05). For further evidence we performed stratified analyses again using FEV1/FVC as the independent variable and CVD as the dependent variable, which showed that the interaction between all variables was not significant (Table S3).

The general population CVD diagnosis ML model

Feature selection showed that the original 12 variables, including age, gender, race, PIR, education, smoking, alcohol, BMI, hyperlipidemia, hypertension, diabetes, and AO, were the optimal set for the machine learning models. Six different ML models, XGBoost, Random Forest, Naive Bayes, CatBoost, KNN, and LightGBM, were used to predict CVD in the general population. The data were divided into a training set (80%) and a testing set (20%) to optimize model training, testing, and evaluate the model's generalization ability. Table 4 shows the performance metrics for all six ML models.

Among these models, CatBoost and XGBoost showed the best performance, with CatBoost achieving the highest AUC on the test set (AUC: 0.7346) and XGBoost achieving a slightly lower but still excellent AUC (AUC: 0.7508). Figure 2A shows the AUC values and receiver operating characteristic (ROC) curves for all six ML models. Figure 2B displays the PR curves (Precision-Recall Curve) of the six ML models, demonstrating the relationship between precision and recall at different

Table 1 Baseline characteristics of all participants

Characteristic	CVD				
	N[1]	Overall, *N* = 12,052 (100%)[2]	No, *N* = 10,406 (89%)[2]	Yes, *N* = 1,646 (11%)[2]	*P*Value[3]
Age	12,052	54.0 (46.0, 64.0)	53.0 (46.0, 63.0)	64.0 (54.0, 72.0)	< 0.001
Age group	12,052				< 0.001
40–59 years		6,296 (64%)	5,812 (68%)	484 (38%)	
60–79 years		5,071 (33%)	4,112 (31%)	959 (57%)	
80 + years		685 (2.1%)	482 (1.7%)	203 (5.7%)	
Gender	12,052				< 0.001
Male		5,806 (48%)	4,926 (47%)	880 (55%)	
Female		6,246 (52%)	5,480 (53%)	766 (45%)	
Race	12,052				0.2
Non-Hispanic white		6,116 (79%)	5,214 (79%)	902 (78%)	
Non-Hispanic black		2,528 (9.0%)	2,161 (8.7%)	367 (11%)	
Mexican-American		2,251 (4.6%)	1,971 (4.7%)	280 (3.8%)	
Other		1,157 (7.7%)	1,060 (7.8%)	97 (7.3%)	
PIR	12,052	3.30 (1.84, 5.00)	3.40 (1.95, 5.00)	2.22 (1.24, 3.97)	< 0.001
PIR group	12,052				< 0.001
Low		3,336 (15%)	2,716 (14%)	620 (27%)	
Medium		4,974 (38%)	4,268 (37%)	706 (44%)	
High		3,742 (47%)	3,422 (49%)	320 (30%)	
Education	12,052				< 0.001
Low		855 (3.2%)	721 (3.0%)	134 (4.6%)	
Middle		5,381 (33%)	4,442 (31%)	939 (47%)	
High		5,816 (64%)	5,243 (66%)	573 (49%)	
Smoking	12,052				< 0.001
Yes		6,343 (53%)	5,336 (52%)	1,007 (65%)	
No		5,709 (47%)	5,070 (48%)	639 (35%)	
Alcohol	12,052				0.2
Never		1,899 (11%)	1,607 (11%)	292 (13%)	
Moderate		9,741 (85%)	8,432 (85%)	1,309 (84%)	
Over		412 (3.6%)	367 (3.6%)	45 (3.2%)	
BMI	12,052	27.50 (24.39, 31.56)	27.38 (24.30, 31.31)	28.60 (25.27, 33.01)	< 0.001
BMI group	12,052				< 0.001
Underweight (< 18.5)		114 (0.9%)	96 (1.0%)	18 (0.8%)	
Normal (18.5 to < 25)		3,141 (29%)	2,759 (29%)	382 (23%)	
Overweight (Overweight (25 to < 30)		4,542 (37%)	3,946 (37%)	596 (33%)	
Obese (≥ 30)		4,255 (34%)	3,605 (33%)	650 (43%)	
Hyperlipidemia	12,052				0.9
No		1,480 (17%)	1,289 (17%)	191 (17%)	
Yes		10,572 (83%)	9,117 (83%)	1,455 (83%)	
Hypertension	12,052				< 0.001
No		5,870 (55%)	5,420 (58%)	450 (32%)	
Yes		6,182 (45%)	4,986 (42%)	1,196 (68%)	
Diabetes	12,052				< 0.001
No		9,661 (85%)	8,568 (87%)	1,093 (71%)	
Yes		2,391 (15%)	1,838 (13%)	553 (29%)	
FEV1	12,052	2,816 (2,242, 3,449)	2,862 (2,289, 3,487)	2,387 (1,858, 3,061)	< 0.001
FVC	12,052	3,713 (3,010, 4,565)	3,757 (3,061, 4,603)	3,314 (2,599, 4,183)	< 0.001
FEV1/FVC	12,052	0.7645 (0.7117, 0.8097)	0.7671 (0.7151, 0.8114)	0.7414 (0.6712, 0.7906)	< 0.001
Airflow obstruction	12,052				< 0.001
No		9,439 (78%)	8,309 (80%)	1,130 (66%)	
Yes		2,613 (22%)	2,097 (20%)	516 (34%)	

[1]N not Missing (unweighted)

Table 1 (continued)

Characteristic	CVD			
	N[1]	Overall, *N* = 12,052 (100%)[2]	No, *N* = 10,406 (89%)[2]	Yes, *N* = 1,646 (11%)[2]	*P*Value[3]

[2]median (IQR) for continuous; n (%) for categorical

[3]Wilcoxon rank-sum test for complex survey samples; chi-squared test with Rao & Scott's second-order correction

Table 2 Association between AO and CVD in all participants (logistic regression model)

Characteristic	Model 1		Model 2		Model 3	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Airflow obstruction		< 0.001		< 0.001		0.003
No	Ref		Ref		Ref	
Yes	2.08 (1.79, 2.42)		1.38 (1.15, 1.65)		1.33 (1.10, 1.61)	

Abbreviations: OR, Odds Ratio; CI, Confidence Interval; Ref, reference

classification thresholds. XGBoost (AP: 0.3186) had the highest average precision (AP) value, indicating that it can achieve relatively high recall while maintaining a certain precision rate when predicting CVD risk. XGBoost showed the most balanced overall performance, and thus was selected as the final prediction model for SHAP visualisation analysis.

The AO population CVD diagnosis ML model

Feature selection validated that the initial 11 variables, including age, gender, race, PIR, education, smoking, alcohol, BMI, hyperlipidemia, hypertension, and diabetes, were the most effective for the machine learning models. Six different ML models, XGBoost, Random Forest, Naive Bayes, LightGBM, KNN, and MLP, were used to identify CVD in the AO population. The data were divided into a training set (80%) and a testing set (20%) to optimise model training, testing, and evaluate the model's generalisation ability. Figure 2C shows the AUC values and receiver operating characteristic (ROC) curves for all six ML models. XGBoost (AUC: 0.6645) and Naive Bayes (AUC: 0.6650) showed excellent AUC performance, implying good classification performance. Figure 2D shows the PR curves (Precision-Recall Curve) of the six ML models, demonstrating the relationship between precision and recall at different classification thresholds. XGBoost (AP: 0.3545) had the highest average precision (AP) value, indicating that it can achieve relatively high recall while maintaining a certain precision rate when predicting CVD risk. The accuracy, Brier Score, F1 Score, Precision, and Recall of all ML models are shown in Table 4. Among the six ML models, XGBoost performs the most balanced performance. Therefore, XGBoost is selected as the final prediction model and analysed for SHAP visualisation.

Visualization of feature importance and personalized predictions

In this study, the XGBoost model utilises SHAP for visualising its results, revealing that all variables significantly impact the model. Specifically, Fig. 3A highlights Age as the most crucial characteristic, altering the predicted absolute probability of CVD risk by roughly 50% on average, whereas Fig. 3B emphasises Education with a similar impact. Figure 4 further illustrates these findings, with each point representing a variable's data set entry distributed horizontally by its SHAP value. The point's colour signifies the feature's value, with red indicating high and blue indicating low. Vertically, the feature's contribution is depicted, increasing with value. Notably, Fig. 4A shows Age contributing most to the eigenvalue, followed by Hypertension and PIR, with age positively correlating with CVD risk. Figure 4B indicates Education Level as the primary contributor, followed by Gender and Race, and suggests that higher education is associated with a lower predicted CVD risk.

In Fig. 5, decision trees are illustrated with a central vertical line representing the base value of the models. Coloured lines indicate individual feature predictions, showing their impact on shifting the output above or below the average prediction, listed in order of decreasing importance. In Fig. 5, the SHAP values accumulate from the base to the final model score, with lines in 5 A converging at -0.0010 and those in 5B at -0.0046. Figures S3A and S3B present CVD risk scores of -0.85 and -3.68, respectively.

Discussion

This study used data from NHANES III and NHANES 2007–2012. Logistic regression showed a significant positive correlation between AO and CVD. We then developed a CVD risk prediction model for the general population, incorporating AO as a predictive factor, and another model specifically for the AO population. XGBoost was selected as the optimal model for both.

Table 3 Subgroup analysis of the association of AO and CVD in all participants

	OR (95% CI)	*P* value	*P* for interaction
Age			0.082
No airflow obstruction			
40–59 years	Ref		
60–79 years	2.78(2.21,3.49)	< 0.0001	
80 + years	4.65(3.09,7.01)	< 0.0001	
Airflow obstruction			
40–59 years	1.87(1.28,2.72)	0.0018	
60–79 years	3.39(2.62,4.40)	< 0.0001	
80 + years	5.74(3.97,8.31)	< 0.0001	
Gender			0.812
No airflow obstruction			
Male	Ref		
Female	0.73(0.60,0.88)	0.0014	
Airflow obstruction			
Male	1.45(1.12,1.88)	0.0068	
Female	1.00(0.74,1.36)	0.9752	
Race			0.275
No airflow obstruction			
Non-Hispanic white	Ref		
Non-Hispanic black	1.07(0.88,1.31)	0.504	
Mexican-American	0.74(0.57,0.96)	0.028	
Other	1.09(0.69,1.74)	0.7022	
Airflow obstruction			
Non-Hispanic white	1.50(1.21,1.86)	0.0004	
Non-Hispanic black	1.15(0.84,1.56)	0.3789	
Mexican-American	1.17(0.78,1.76)	0.4484	
Other	1.02(0.46,2.28)	0.9589	
PIR			0.006
No airflow obstruction			
Low	Ref		
Medium	0.70(0.55,0.89)	0.004	
High	0.38(0.29,0.48)	< 0.0001	
Airflow obstruction			
Low	1.34(1.02,1.77)	0.0403	
Medium	0.72(0.53,1.00)	0.0516	
High	0.82(0.59,1.14)	0.2405	
Education			0.123
No airflow obstruction			
Low	Ref		
Middle	1.46(1.10,1.93)	0.0098	
High	0.93(0.68,1.26)	0.6262	
Airflow obstruction			
Low	2.01(1.12,3.61)	0.0221	
Middle	1.75(1.20,2.57)	0.0051	
High	1.49(1.06,2.09)	0.0252	
BMI			0.524
No airflow obstruction			
Underweight (< 18.5)	Ref		
Normal (18.5 to < 25)	2.02(0.68,5.96)	0.2072	
Overweight (Overweight (25 to < 30)	2.05(0.70,5.96)	0.193	
Obese (≥ 30)	3.00(1.02,8.80)	0.0491	
Airflow obstruction			
Underweight (< 18.5)	3.33(0.73,15.20)	0.125	

Table 3 (continued)

	OR (95% CI)	*P* value	*P* for interaction
Normal (18.5 to < 25)	2.77(0.92,8.30)	0.0735	
Overweight (Overweight (25 to < 30)	3.27(1.07,10.00)	0.0408	
Obese (≥ 30)	3.80(1.28,11.26)	0.0183	
Smoking			0.444
No airflow obstruction			
Yes	Ref		
No	0.66(0.54,0.81)	0.0002	
Airflow obstruction			
Yes	1.49(1.20,1.84)	0.0005	
No	0.83(0.57,1.21)	0.3315	
Alcohol			0.129
No airflow obstruction			
Never	Ref		
Moderate	0.85(0.67,1.08)	0.1867	
Over	0.56(0.33,0.95)	0.0359	
Airflow obstruction			
Never	0.89(0.57,1.37)	0.5858	
Moderate	1.24(0.94,1.64)	0.1288	
Over	1.30(0.57,2.94)	0.5384	
Hyperlipidemia			0.504
No airflow obstruction			
Yes	Ref		
No	0.94(0.68,1.31)	0.7295	
Airflow obstruction			
Yes	1.66(1.00,2.75)	0.0543	
No	1.30(0.92,1.82)	0.1408	
Hypertension			0.908
No airflow obstruction			
Yes	Ref		
No	1.84(1.46,2.32)	< 0.0001	
Airflow obstruction			
Yes	1.44(1.05,1.97)	0.0245	
No	2.60(1.94,3.49)	< 0.0001	
Diabetes			0.579
No airflow obstruction			
Yes	Ref		
No	1.83(1.45,2.32)	< 0.0001	
Airflow obstruction			
Yes	1.46(1.16,1.84)	0.0017	
No	2.37(1.69,3.33)	< 0.0001	

Abbreviations: OR, Odds Ratio; CI, Confidence Interval; Ref, reference

Age, hypertension, and PIR have the greatest impact on CVD risk prediction in the general population. For the AO population, education, gender, and race are the most influential factors in CVD risk prediction.

AO is closely related to CVD. Multiple studies have indicated that reduced lung function indicators, such as FEV1% pred, FVC% pred, and FEV1/FVC, are significantly associated with a high risk of CVD prevalence and mortality, which aligns with our research findings [15, 17, 43]. A 5% reduction in FEV1/FVC is associated with a 0.47% increase in the 10-year risk of CVD ($P < 0.001$) [44]. Improving FEV1 can effectively reduce the risk of CVD mortality [43]. AO is a crucial manifestation of decreased lung function, with reversible AO being a sign of asthma and irreversible AO primarily indicative of COPD [45]. COPD and asthma often coexist with CVD [46, 47]. Compared to patients without COPD, those with COPD have a higher risk of ischemic heart disease, HF, arrhythmias, or peripheral vascular diseases [48]. Comorbidities lead to an increased risk of hospitalization and poorer prognosis for patients [49, 50]. According to statistics, CVD accounts for 42% of the first hospitalization and

Table 4 Comparisons of six machine learning classifiers

	Accuracy		Precision		Recall		F1 Score		Brier Score	
	Train set	Test set	Train set	Test set	Train set	Test set	Train set	Test set	Train set	Test set
Full Population CVD Diagnostic Model										
XGBoost	0.7157	0.6665	0.6969	0.2524	0.7634	0.7062	0.7286	0.3719	0.1911	0.2012
Random Forest	0.6928	0.6433	0.6709	0.2408	0.7570	0.7211	0.7113	0.3611	0.2067	0.2110
KNN	1.0000	0.7350	1.0000	0.2245	1.0000	0.3650	1.0000	0.2780	0. 0000	0.2172
Naive Bayes	0.6709	0.6591	0.6650	0.2508	0.6888	0.7240	0.6767	0.3725	0. 2118	0.2107
Catboost	0.7581	0.6765	0.7314	0.2525	0.8159	0.6706	0.7713	0.3669	0.1720	0.1950
LightGBM	0.7166	0.6653	0.6967	0.2495	0.7674	0.6944	0.7303	0.3671	0.1883	0.2013
Airflow Obstruction Population CVD Diagnostic Model										
XGBoost	0.7671	0.6539	0.7521	0.3085	0.7968	0.5321	0.7738	0.3906	0.1608	0.2162
Random Forest	0.9831	0.6902	0.9672	0.2977	1.000	0.3578	0.9833	0.3250	0.0189	0.1983
MLP	0.9468	0.6539	0.9367	0.2750	0.9584	0.4037	0.9474	0.3271	0.0417	0.2930
KNN	1.0000	0.6252	1.0000	0.2570	1.0000	0.4220	1.0000	0.3194	0.0000	0.2840
Naive Bayes	0.6786	0.5698	0.6368	0.2929	0.8313	0.7523	0.7211	0.4216	0.2106	0.2643
LightGBM	0.7748	0.6444	0.7582	0.2941	0.8069	0.5046	0.7818	0.3716	0.1587	0.2189

Abbreviations: XGBoost, eXtreme Gradient Boosting; LightGBM, Light Gradient Boosting Machine; KNN, K-Nearest Neighbor; MLP, Multi-Layer Perceptron; CatBoost, Categorical Boosting

44% of the second hospitalization among COPD patients, suggesting a possible association between reduced FEV1 and increased mortality caused by cardiovascular complications induced by COPD [51]. Approximately 50% of COPD deaths are attributed to cardiovascular events, particularly congestive HF, arrhythmias, and acute myocardial infarction [52–55]. Compared to the general population, asthma patients have a higher risk and prevalence of CVD. Patients with early-onset asthma have a 26% higher risk of CVD than those without asthma, while patients with late-onset asthma have a 39% increased risk of CVD [56]. These findings reinforce the strong link between AO and CVD, suggesting that we should fully consider the impact of lung function and early diagnose and identify high-risk groups for CVD.

AO can affect the cardiovascular system through various mechanisms. There exists a close pathophysiological link between COPD and CVD, as factors such as systemic inflammation, oxidative stress, hypoxia, and hyperinflation exacerbate the progression of CVD [57]. Oxidative stress and increased systemic inflammation caused by COPD can lead to vascular remodelling and arterial stiffness, resulting in atherosclerosis and altering the vascular structure [58]. This includes thickening of the arterial wall, increasing the likelihood of atheromatous plaque or lesion formation, which can lead to myocardial infarction or stroke. Simultaneously, it also causes changes and structural remodelling in cerebral blood vessels, promoting the disruption of the blood-brain barrier [59, 60]. Hypoxia can also cause vascular remodelling and endothelial dysfunction, increasing vascular resistance and inducing vasoconstriction, thus aggravating the risk of pulmonary hypertension [61, 62]. Furthermore, hyperinflation is associated with impaired ventricular filling, reduced cardiac output, and increased pulmonary vascular resistance, adding to the burden on the heart [63, 64]. A similar association exists between asthma and CVD. Both asthma and CVD are related to chronic inflammation. The 5-lipo-oxygenase (5-LOX) pathway is a common mechanism in both asthma and CVD. Asthma can lead to overactivation of the 5-LOX pathway, resulting in the production of large amounts of leukotrienes, which stimulate vascular smooth muscle cell proliferation, migration, and inflammatory responses, promoting the formation and development of atherosclerosis [65]. There is also overlap in inflammatory mediators between asthma and CVD patients. Common systemic inflammatory markers, such as IL-6, C-reactive protein, fibrinogen, and D-dimer, are increased in both asthma and CVD [66, 67]. Specifically, IL-6 and TNF-α, two common proinflammatory factors in asthma, have been shown to have a close relationship with atherosclerosis [68]. These inflammatory mediators not only play a crucial role in the pathogenesis of asthma and CVD, but they may also exacerbate the progression and aggravation of each other.

Shared risk factors play a crucial role in explaining the coexistence of cardiopulmonary comorbidities and AO-related diseases [49, 69, 70]. Aging, hypertension, diabetes, and smoking all contribute to an increased risk of CVD and are more prevalent in COPD. As individuals age, structural and functional changes in the respiratory and cardiovascular systems may occur. The elderly population is more susceptible to AO symptoms such as COPD and is also at higher risk for developing CVD [71]. Smoking is widely recognized as a primary risk factor for both AO and CVD. Harmful substances in smoke can damage endothelial cells in the respiratory tract and blood vessels, leading to inflammatory responses and oxidative stress that promote the occurrence of AO and CVD [72]. Hypertension increases cardiac workload,

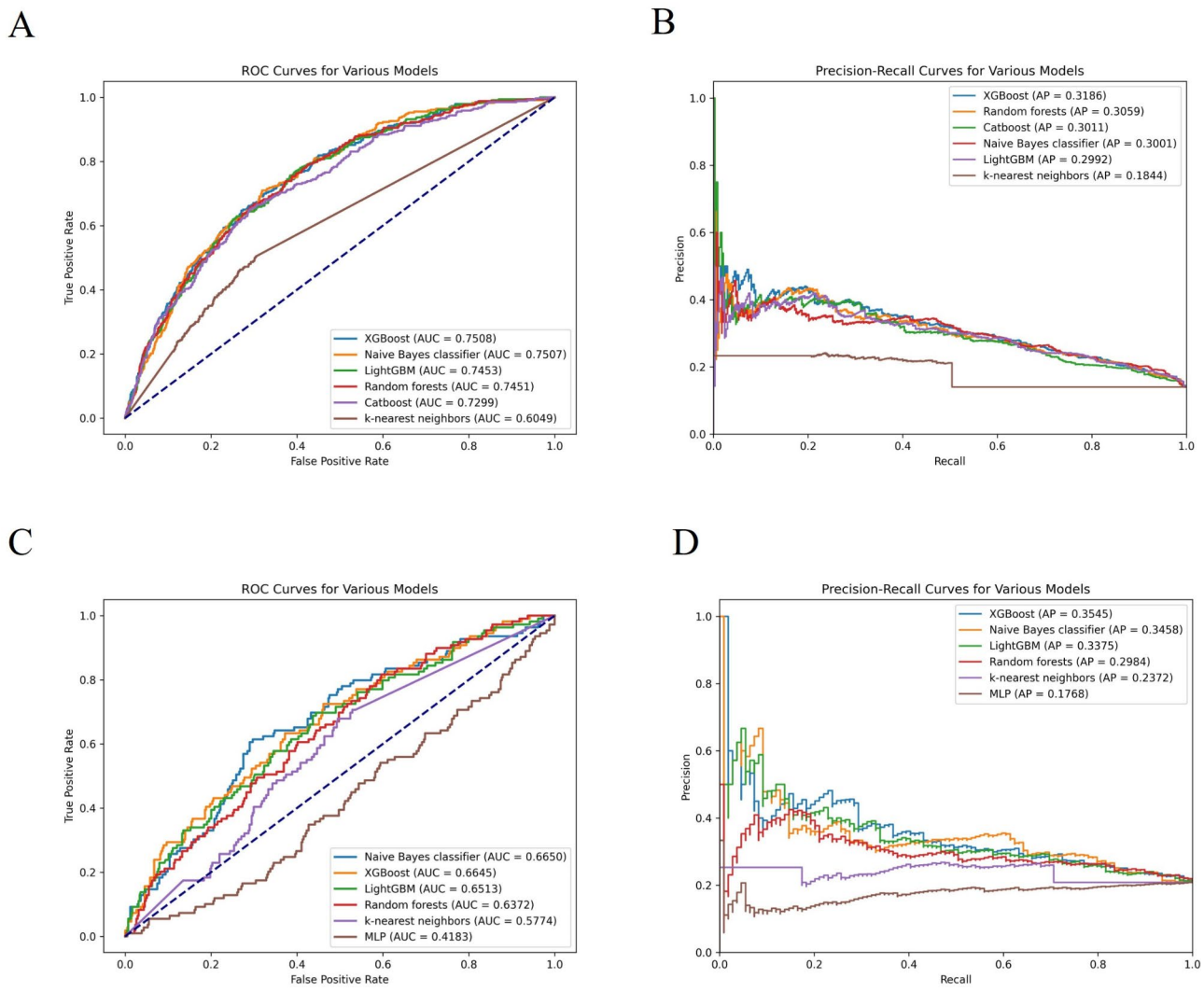

Fig. 2 The ROC curves and PR curves of the six ML models. (**A**) The AUC values and ROC curves for the Full Population Model. (**B**) The AP values and PR Curves for the Full Population Model. (**C**) The AUC values and ROC curves for the Airflow Obstruction Population Model. (**D**) The AP values and PR Curves for the Airflow Obstruction Population Model.

resulting in myocardial hypertrophy and impaired heart function; meanwhile, diabetes affects normal vascular function, increasing the risk of atherosclerosis. These conditions may exacerbate the CVD risk among patients with AO through various mechanisms including influencing inflammatory responses, microcirculation, neuroregulation, cardiac load, and atherosclerosis [73]. This aligns with our study findings, where SHAP analysis indicated that age and hypertension were the two most significant features in predicting CVD in the general population.

However, for individuals with airflow obstruction, SHAP analysis identified some differences in the most influential factors compared to previous understanding. Education level, gender, and ethnicity were found to have the greatest impact on predicting CVD risk. Education level, a key indicator of socioeconomic status, is closely associated with CVD risk. Higher education often correlates with greater health knowledge, healthier lifestyles, and better access to healthcare resources, all of which collectively reduce CVD risk. Educated individuals typically have higher health literacy, enabling them to effectively understand health information and make positive lifestyle adjustments to lower their CVD risk [74–76]. They are also more likely to accurately report their health status, enhancing the influence of education in statistical models. While the impact of education is significant, SHAP analysis also highlighted the importance of gender and ethnicity. This reflects the complex interplay of social determinants of health. Gender and ethnicity affect access to healthcare, exposure to environmental risks, and socioeconomic status, all of which can influence CVD risk [77, 78]. Similarly, gender differences in

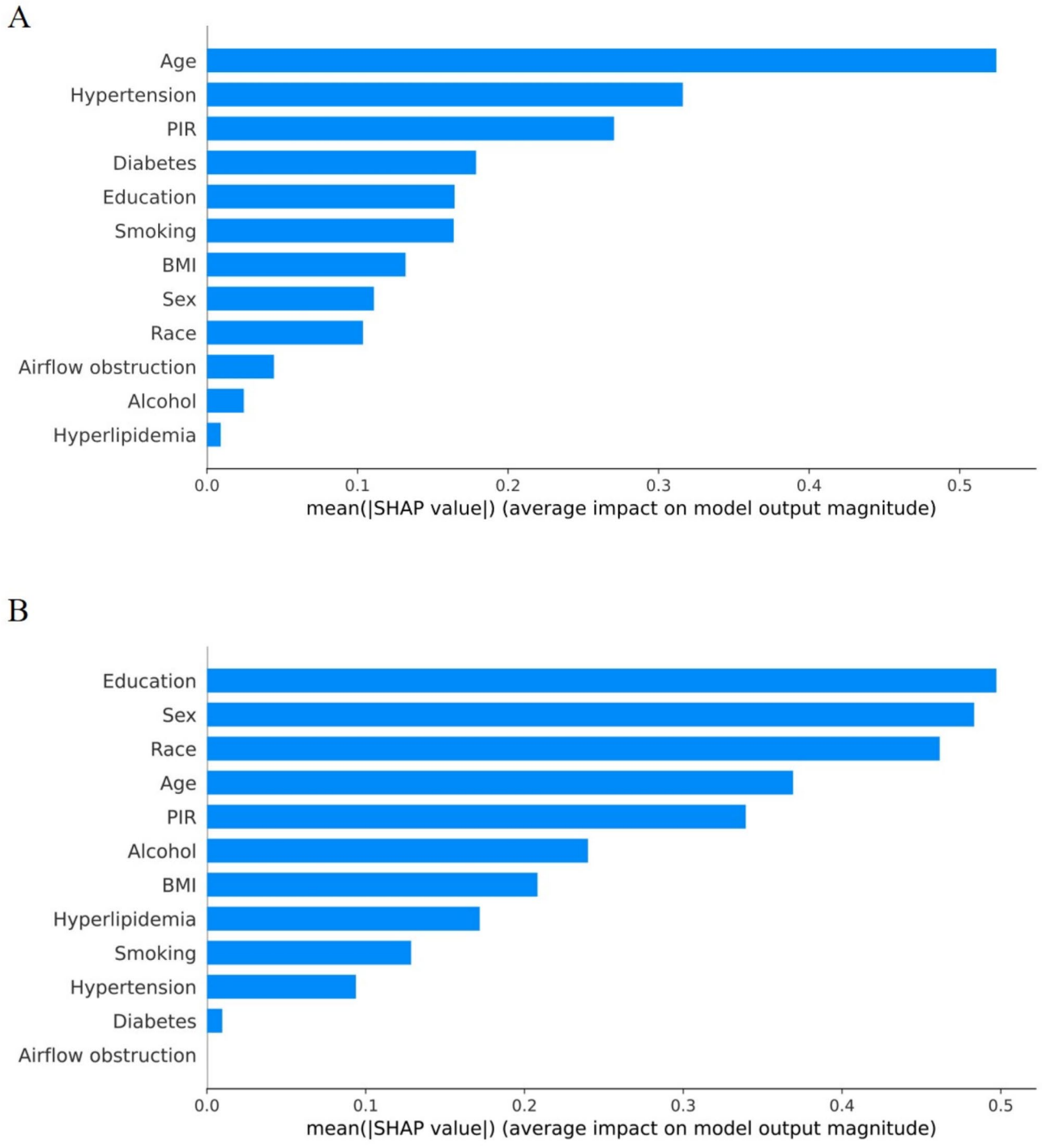

Fig. 3 The importance scores of features for the XGBoost model. (**A**) Feature Importance Scores for the Full Population CVD Diagnostic Model. (**B**) Feature Importance Scores for the Airflow Obstruction Population CVD Diagnostic Model.

healthcare utilisation and social roles can also impact CVD outcomes.

In this study, the developed machine learning model, particularly XGBoost, has shown effectiveness in predicting CVD among the general population and a specific AO population. For the general population, the AUC value of XGBoost is 0.7508; for the AO population, it reaches 0.6645, demonstrating performance comparable to previous studies focused on CVD prediction. The models tailored for specific populations indicate that ML can improve CVD risk stratification and detection. In a multiethnic patient group, the AUC of XGBoost algorithm in

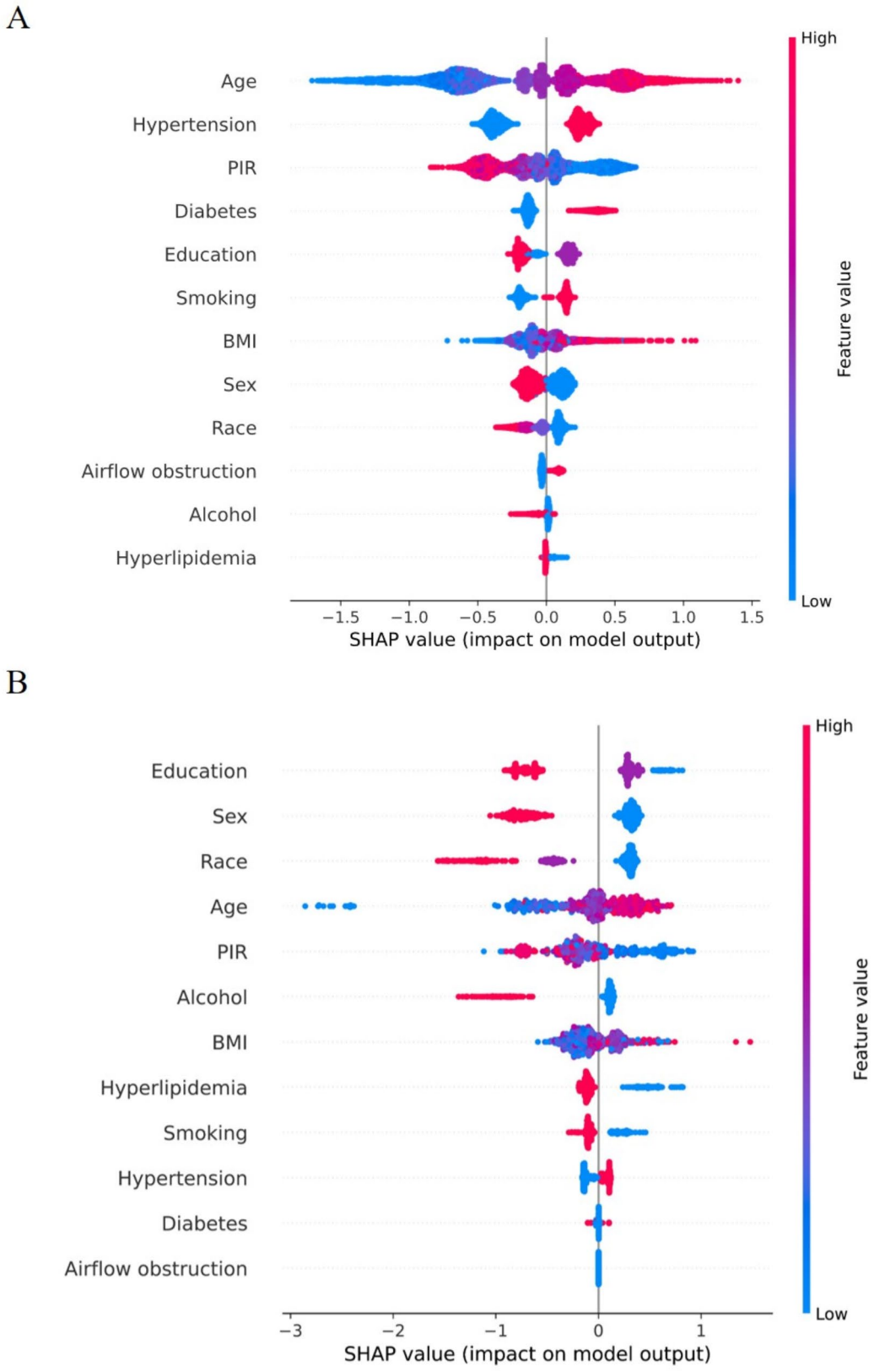

Fig. 4 The SHAP summary plot of the XGBoost model. (**A**) The SHAP summary plot for the Full Population CVD Diagnostic Model. (**B**) The SHAP summary plot for the Airflow Obstruction Population CVD Diagnostic Model.

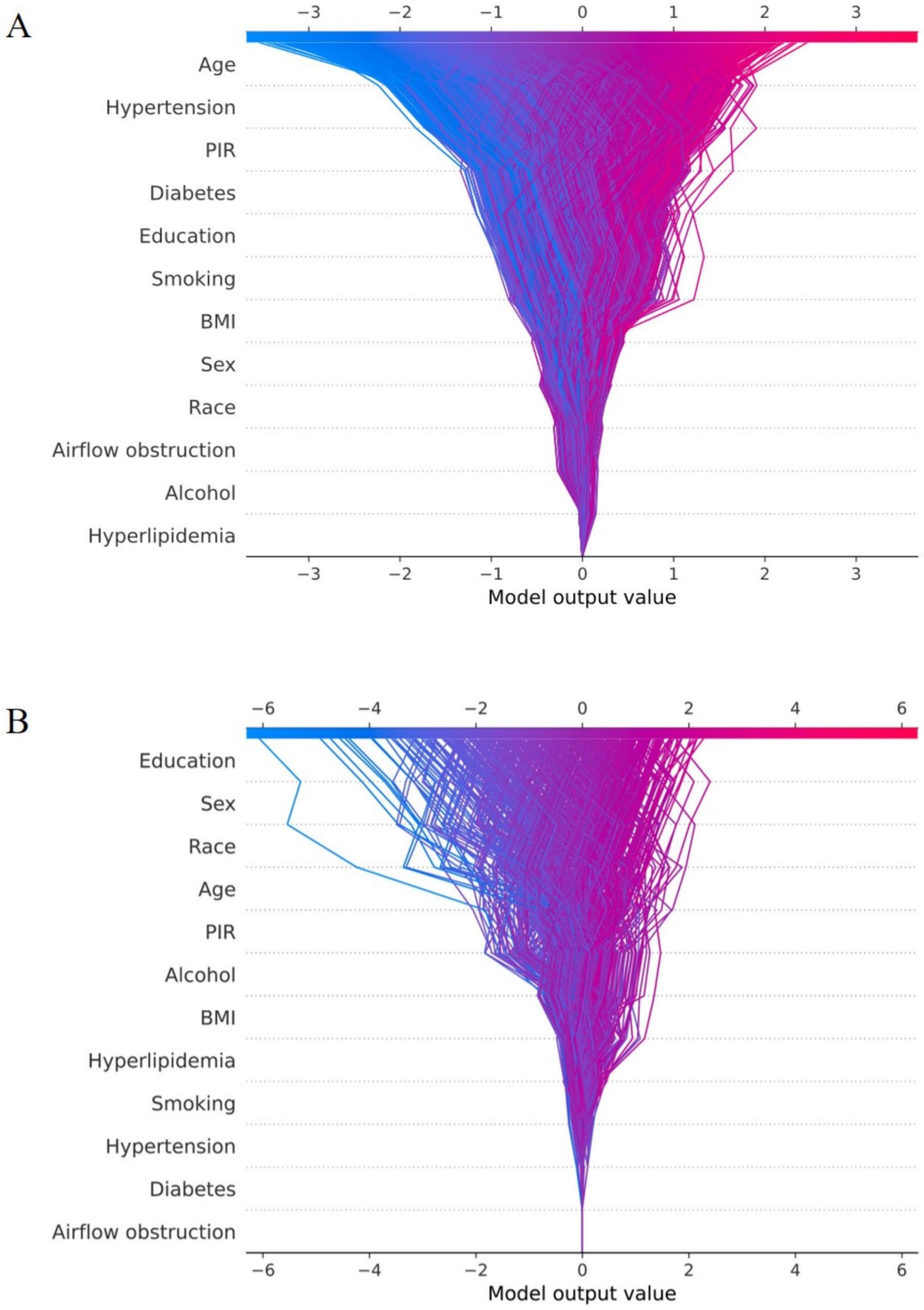

Fig. 5 The SHAP decision plot of the XGBoost model. (**A**) The SHAP decision plot for the Full Population CVD Diagnostic Model. (**B**) The SHAP decision plot for the Airflow Obstruction Population CVD Diagnostic Model.

predicting CVD events within five years ranges from 0.70 to 0.71 [79]. For specific ethnic groups, the AUC range for CVD prevalence detection by ML models is 0.660 to 0.74 [80]. Additionally, a random forest model specifically developed for type 2 diabetes patients achieved an AUC of 0.722 in predicting CVD development [81]. However, compared to CVD risk prediction models based on the general population, the performance of the XGBoost model in this study is slightly inferior, with AUC values marginally lower than those in existing research [82]. This may be due to the XGBoost model in this study placing greater emphasis on data interpretability and clinical applicability during its construction, at the cost of some predictive precision. Nonetheless, the XGBoost model in this study retains significant advantages, having developed dedicated models for underrepresented AO populations to enhance the relevance and accuracy of predictions, while adopting more robust algorithms to ensure the stability and reliability of prediction outcomes. In practice, these models assist healthcare providers in identifying high-risk individuals, facilitating early intervention and prevention.

In comparison with previous studies, the present research possesses several advantages. Firstly, existing models for predicting CVD have not taken into account the specific population with AO. To our knowledge, this is the first study to develop a prediction model for CVD risk in patients with AO using a ML approach. This not only deepens our understanding of the potential relationship between AO and CVD, but also promises to bring more precise and personalized treatment and management plans for patients with AO. Secondly, the study employed more rigorous statistical methods and a larger sample size. The data is rich and representative, which is conducive to ensuring the extensiveness and applicability of the research results, thus enhancing the external validity of the study. Moreover, the research did not merely stop at the level of relationship analysis. It further utilized various ML methods to construct a CVD risk prediction model for the AO population. Emerging techniques such as SMOTE, RandomizedSearchCV, and cross-validation were also employed to enhance the robustness and predictive power of the model. This reflects the depth and forward-looking nature of the research.

However, this study has several limitations. Firstly, the cross-sectional design limits our ability to establish a causal relationship between AO and CVD. Longitudinal studies are needed to confirm the temporal relationship between these conditions. The cohort used in this study is derived solely from the NHANES database, which may introduce specific limitations and biases. As non-Hispanic whites dominate the NHANES population, the current findings may not fully represent the realities of other ethnic groups. Additionally, the assessment of CVD was based on patient self-reports, which likely introduces significant bias. Furthermore, constrained by the complex impacts of data availability and computational resources, the performance of our machine learning ensemble method did not meet expectations, and we were unable to propose a new ensemble model. Future longitudinal studies should aim for a larger and more diverse sample size, along with comprehensive data collection, to further investigate the causal relationship between AO and CVD and to refine the CVD risk prediction models.

This research provides valuable insights into the relationship between AO and CVD. The application of ML technology has resulted in a more comprehensive and sophisticated risk assessment tool, which contributes to the early identification and management of CVD in individuals with AO. This alleviates the disease burden of both AO and CVD, and improves the overall health of the affected individuals.

Conclusion

This study demonstrates a significant positive correlation between AO and CVD. Among the general population, age, hypertension, and PIR have the greatest impact on CVD risk prediction. For individuals with AO, education, gender, and ethnicity emerge as the most influential factors in CVD risk prediction. Developing accurate models to predict CVD risk, particularly in the AO population, enhances early identification and intervention of high-risk individuals. This research, by advocating for more comprehensive risk assessment tools, contributes to improving the prevention and management of CVD.

Abbreviations

AO	Airflow obstruction
AP	Average Precision
AUC	Area Under the Curve
CHD	Coronary heart disease
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular Disease
FEV1	Forced expiratory volume in 1 s
FVC	Forced vital capacity
HF	Heart failure
KNN	K-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine
ML	Machine learning
MLP	Multilayer Perceptron
CatBoost	Categorical Boosting
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Survey
PIR	Ratio of family income to poverty
ROC	Receiver operating characteristic
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-Sampling Technique
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
VIF	Variance inflation factor
XGBoost	eXtreme Gradient Boosting

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02885-0 .

Supplementary Material 1. Angina defined by positive results of ROSE-questionnaire in NHANES III

Supplementary Material 2. The VIF values of variables

Supplementary Material 3. TheSHAP force plot of the XGBoost model. (A) The SHAP force plot for the Full Population CVD Diagnostic Model. (B) The SHAP force plot for the Airflow Obstruction Population CVD Diagnostic Model.

Supplementary Material 4

Author contributions

C. Zhang was primarily responsible for the conceptualization and design of the study. X. Cao, J. Ma collected the data. X. Cao, J. Ma, and X. He contributed to the data analysis and interpretation of the results. X. Cao and J. Ma wrote the initial manuscript. Y. Liu, Y. Yang, and Y. Wang revised the final article. All authors read and approved the final manuscript.

Funding

2022 "Tianfu Qingcheng Plan" Tianfu Science and Technology Leading Talents Project (Chuan Qingcheng No. 1090). The National TCM Clinical Excellent Talents Training Program (National TCM Renjiao Letter [2022] No. 1). Sichuan Science and Technology Program (2023ZYD0050, 2024NSFJQ0059). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability

The datasets generated and analyzed in the current study are available at NHANES website: https://www.cdc.gov/nchs/nhanes/index.htm. The code will be made available on request.

Declarations

Ethics approval and consent to participate

This research entailed the analysis of de-identified information retrieved from the public database of the NHANES. Ethical approval was granted by the NCHS Ethics Review Committee. All methods adhered to the relevant guidelines and regulations, including the Declaration of Helsinki. Additionally, all participants provided written informed consent prior to their involvement in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 June 2024 / Accepted: 20 January 2025
Published online: 03 February 2025

References

1. Sattelmair J, Pertman J, Ding EL, Kohl HW, Haskell W, Lee IM. Dose response between physical activity and risk of Coronary Heart Disease A Meta-Analysis. Circulation. 2011;124(7):789–.
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng SS, Delling FN, et al. Heart Disease and Stroke Statistics-2021 Update A Report from the American Heart Association. Circulation. 2021;143(8):e254–743.
3. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. The physical activity guidelines for americans. JAMA. 2018;320(19):2020–8.
4. Kassebaum NJ, Arora M, Barber RM, Bhutta ZA, Carter A, Casey DC, Charlson FJ, Coates MM, Coggeshall M, Cornaby L, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the global burden of Disease Study 2015. Lancet. 2016;388(10053):1603–58.
5. Omran AR. The epidemiologic transition. A theory of the epidemiology of population change. Milbank Q. 1971;49(4):509–38.
6. Cardiovascular diseases (CVDs.) [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)].
7. Global regional. National life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of Disease Study 2015. Lancet. 2016;388(10053):1459–544.
8. Townsend N, Kazakiewicz D, Wright FL, Timmis A, Huculeci R, Torbica A, Gale CP, Achenbach S, Weidinger F, Vardas P. Epidemiology of cardiovascular disease in Europe. Nat Rev Cardiol. 2022;19(2):133–43.
9. Liu SW, Li YC, Zeng XY, Wang HD, Yin P, Wang LJ, Liu YN, Liu JM, Qi JL, Ran S, et al. Burden of Cardiovascular diseases in China, 1990–2016 findings from the 2016 global burden of Disease Study. JAMA Cardiol. 2019;4(4):342–52.
10. Brown JC, Gerhardt TE, Kwon E. Risk Factors for Coronary Artery Disease. In: *StatPearls*. edn. Treasure Island (FL) ineligible companies. Disclosure: Thomas Gerhardt declares no relevant financial relationships with ineligible companies. Disclosure: Edward Kwon declares no relevant financial relationships with ineligible companies.: StatPearls Publishing Copyright © 2024, StatPearls Publishing LLC.; 2024.
11. Multimorbidity. clinical assessment and management [https://www.nice.org.uk/guidance/ng56]
12. Feng WJ, Zhang ZY, Liu Y, Li ZB, Guo WJ, Huang FF, Zhang JW, Chen AL, Ou CW, Zhang K, et al. Association of Chronic respiratory symptoms with Incident Cardiovascular Disease and all-cause mortality findings from the coronary artery risk development in young adults study. Chest. 2022;161(4):1036–45.
13. Buhr RG, Barjaktarevic IZ, Quibrera PM, Bateman LA, Bleecker ER, Couper DJ, Curtis JL, Dolezal BA, Han MLK, Hansel NN, et al. Reversible airflow obstruction predicts Future Chronic Obstructive Pulmonary Disease Development in the SPIROMICS Cohort An Observational Cohort Study. Am J Respir Crit Care Med. 2022;206(5):554–62.
14. Collaro AJ, Chang ANB, Marchant JM, Chatfield MD, Dent A, Blake T, Mawn P, Fong K, McElrea MS. Associations between lung function and future cardiovascular morbidity and overall mortality in a predominantly first nations population: a cohort study. Lancet Reg Health-W Pac. 2021;13:8.
15. Costanzo S, Magnacca S, Bonaccio M, Di Castelnuovo A, Piraino A, Cerletti C, de Gaetano G, Donati MB, Iacoviello L. Moli-Sani Study I: reduced pulmonary function, low-grade inflammation and increased risk of total and cardiovascular mortality in a general adult population: prospective results from the Moli-Sani study. Respir Med. 2021;184:10.
16. Engström G, Lind P, Hedblad B, Wollmer P, Stavenow L, Janzon L, Lindgärde F. Lung function and cardiovascular risk: relationship with inflammation-sensitive plasma proteins. Circulation. 2002;106(20):2555–60.
17. Min KB, Min JY. Reduced lung function, C-Reactive protein, and increased risk of Cardiovascular Mortality. Circ J. 2014;78(9):2309–U2423.
18. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ (Clinical Res ed). 1996;313(7059):711–5. discussion 715–716.
19. Sin DD, Man SF. Why are patients with chronic obstructive pulmonary disease at increased risk of cardiovascular diseases? The potential role of systemic inflammation in chronic obstructive pulmonary disease. Circulation. 2003;107(11):1514–9.
20. Agustí A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. Lancet Resp Med. 2017;5(12):935–45.
21. Schroeder EB, Welch VL, Couper D, Nieto FJ, Liao D, Rosamond WD, Heiss G. Lung function and incident coronary heart disease: the atherosclerosis risk in communities Study. Am J Epidemiol. 2003;158(12):1171–81.
22. Kim JJ, Kim DB, Jang SW, Cho EJ, Chang K, Baek SH, Youn HJ, Chung WS, Seung KB, Rho TH, et al. Relationship between airflow obstruction and coronary atherosclerosis in asymptomatic individuals: evaluation by coronary CT angiography. Int J Cardiovasc Imaging. 2018;34(4):641–8.
23. Hozawa A, Billings JL, Shahar E, Ohira T, Rosamond WD, Folsom AR. Lung function and ischemic stroke incidence: the atherosclerosis risk in communities study. Chest. 2006;130(6):1642–9.
24. Lee HM, Liu MA, Barrett-Connor E, Wong ND. Association of lung function with coronary heart disease and cardiovascular disease outcomes in elderly: the Rancho Bernardo study. Respir Med. 2014;108(12):1779–85.

25. Wannamethee SG, Shaper AG, Rumley A, Sattar N, Whincup PH, Thomas MC, Lowe GD. Lung function and risk of type 2 diabetes and fatal and nonfatal major coronary heart disease events: possible associations with inflammation. Diabetes Care. 2010;33(9):1990–6.
26. Willerson JT, Ridker PM. Inflammation as a cardiovascular risk factor. Circulation. 2004;109(21 Suppl 1):Ii2–10.
27. Corlateanu A, Covantev S, Mathioudakis AG, Botnaru V, Cazzola M, Siafakas N. Chronic obstructive Pulmonary Disease and Stroke. COPD-J Chronic Obstr Pulm Dis. 2018;15(4):405–13.
28. Maclay JD, MacNee W. Cardiovascular Disease in COPD mechanisms. Chest. 2013;143(3):798–807.
29. Sabia S, Shipley M, Elbaz A, Marmot M, Kivimaki M, Kauffmann F, Singh-Manoux A. Why does lung function predict mortality? Results from the Whitehall II Cohort Study. Am J Epidemiol. 2010;172(12):1415–23.
30. Lange P, Mogelvang R, Marott JL, Vestbo J, Jensen JS. Cardiovascular morbidity in COPD: a study of the general population. Copd. 2010;7(1):5–10.
31. Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. Chest. 2005;127(6):1952–9.
32. Xia B, Innab N, Kandasamy V, Ahmadian A, Ferrara M. Intelligent cardiovascular disease diagnosis using deep learning enhanced neural network with ant colony optimization. Sci Rep. 2024;14(1):16.
33. Asadi F, Homayounfar R, Mehrali Y, Masci C, Talebi S, Zayeri F. Detection of cardiovascular disease cases using advanced tree-based machine learning algorithms. Sci Rep. 2024;14(1):22230.
34. Rifin HM, Omar MA, Wan KS, Hasani WSR. 10-year risk for cardiovascular diseases according to the WHO prediction chart: findings from the National Health and Morbidity Survey (NHMS) 2019. BMC Public Health. 2024;24(1):8.
35. van Apeldoorn JAN, Hageman SHJ, Harskamp RE, Agyemang C, van den Born BJH, van Dalen JW, Galenkamp H, Hoevenaar-Blom MP, Richard E, van Valkengoed IGM, et al. Adding ethnicity to cardiovascular risk prediction: external validation and model updating of SCORE2 using data from the HELIUS population cohort. Int J Cardiol. 2024;417:7.
36. Zheng DZ, Cai JM, Xu SF, Jiang SY, Li CL, Wang B. The association of triglyceride-glucose index and combined obesity indicators with chest pain and risk of cardiovascular disease in American population with pre-diabetes or diabetes. Front Endocrinol. 2024;15:13.
37. Dorraki M, Liao Z, Abbott D, Psaltis PJ, Baker E, Bidargaddi N, Wardill HR, van den Hengel A, Narula J, Verjans JW. Improving Cardiovascular Disease Prediction with Machine Learning using Mental Health data: a prospective UK Biobank Study. JACC Adv. 2024;3(9):101180.
38. Gardner RM. Standardization of spirometry: a summary of recommendations from the American Thoracic Society. The 1987 update. Ann Intern Med. 1988;108(2):217–20.
39. Higbee DH, Granell R, Sanderson E, Smith GD, Dodd JW. Lung function and cardiovascular disease: a two-sample mendelian randomisation study. Eur Resp J. 2021;58(3):7.
40. Sontrop JM, Dixon SN, Garg AX, Buendia-Jimenez I, Dohein O, Huang SH, Clark WF. Association between water intake, chronic kidney disease, and cardiovascular disease: a cross-sectional analysis of NHANES data. Am J Nephrol. 2013;37(5):434–42.
41. Will JC, Yuan K, Ford E. National trends in the prevalence and medical history of angina: 1988 to 2012. Circulation Cardiovasc Qual Outcomes. 2014;7(3):407–13.
42. Huang N, Tang C, Li S, Ma W, Zhai X, Liu K, Sheerah HA, Cao J. Association of lung function with the risk of cardiovascular diseases and all-cause mortality in patients with diabetes: results from NHANES III 1988–1994. Front Cardiovasc Med. 2022;9:976817.
43. Ching SM, Chia YC, Lentjes MAH, Luben R, Wareham N, Khaw KT. FEV1 and total Cardiovascular mortality and morbidity over an 18 years follow-up Population-based prospective EPIC-NORFOLK Study. BMC Public Health. 2019;19:10.
44. Wang B, Zhou Y, Xiao LL, Guo YJ, Ma JX, Zhou M, Shi TM, Tan AJ, Yuan J, Chen WH. Association of lung function with cardiovascular risk: a cohort study. Respir Res. 2018;19:9.
45. Rogliani P, Ora J, Puxeddu E, Cazzola M. Airflow obstruction: is it asthma or is it COPD? Int J Chronic Obstr Pulm Dis. 2016;11:3007–13.
46. Morgan AD, Zakeri R, Quint JK. Defining the relationship between COPD and CVD: what are the implications for clinical practice? Ther Adv Respir Dis. 2018;12:16.
47. Hawkins NM, Petrie MC, Jhund PS, Chalmers GW, Dunn FG, McMurray JJ. Heart failure and chronic obstructive pulmonary disease: diagnostic pitfalls and epidemiology. Eur J Heart Fail. 2009;11(2):130–9.
48. Chen W, Thomas J, Sadatsafavi M, FitzGerald JM. Risk of cardiovascular comorbidity in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Lancet Respiratory Med. 2015;3(8):631–9.
49. Balbirsingh V, Mohammed AS, Turner AM, Newnham M. Cardiovascular disease in chronic obstructive pulmonary disease: a narrative review. Thorax. 2022;77(9):939–45.
50. Gulsvik A, Hansteen V, Sivertssen E. Cardiac arrhythmias in patients with serious pulmonary diseases. Scandinavian J Respiratory Dis. 1978;59(3):154–9.
51. Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. Eur Respir J. 2008;32(4):962–9.
52. Sin DD, Man SF. Chronic obstructive pulmonary disease as a risk factor for cardiovascular morbidity and mortality. Proc Am Thorac Soc. 2005;2(1):8–11.
53. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society 2008;5(4):549–555.
54. Berger JS, Sanborn TA, Sherman W, Brown DL. Effect of chronic obstructive pulmonary disease on survival of patients with coronary heart disease having percutaneous coronary intervention. Am J Cardiol. 2004;94(5):649–51.
55. Curkendall SM, DeLuise C, Jones JK, Lanes S, Stang MR, Goehring E Jr., She D. Cardiovascular disease in patients with chronic obstructive pulmonary disease, Saskatchewan Canada cardiovascular disease in COPD patients. Ann Epidemiol. 2006;16(1):63–70.
56. Zhang B, Li ZF, An ZY, Zhang L, Wang JY, Hao MD, Jin YJ, Li D, Song AJ, Ren Q, et al. Association between Asthma and all-cause Mortality and Cardiovascular Disease Morbidity and Mortality: a Meta-analysis of Cohort studies. Front Cardiovasc Med. 2022;9:14.
57. de Miguel-Díez J, Núñez Villota J, Santos Pérez S, Manito Lorite N, Alcázar Navarrete B, Delgado Jiménez JF, Soler-Cataluña JJ, Pascual Figal D, Sobradillo Ecenarro P. Gómez Doblas JJ: Multidisciplinary Management of patients with Chronic Obstructive Pulmonary Disease and Cardiovascular Disease. Arch Bronconeumol. 2024;60(4):226–37.
58. Luo M, Zheng Y, Tang S, Gu L, Zhu Y, Ying R, Liu Y, Ma J, Guo R, Gao P, et al. Radical oxygen species: an important breakthrough point for botanical drugs to regulate oxidative stress and treat the disorder of glycolipid metabolism. Front Pharmacol. 2023;14:1166178.
59. Brassington K, Selemidis S, Bozinovski S, Vlahos R. New frontiers in the treatment of comorbid cardiovascular disease in chronic obstructive pulmonary disease. Clin Sci. 2019;133(7):885–904.
60. Engin A. Endothelial Dysfunction in Obesity. In: Obesity and Lipotoxicity. Volume 960, edn. Edited by Engin AB, Engin A. Cham: Springer International Publishing Ag; 2017: 345–379.
61. Vivodtzev I, Tamisier R, Baguet JP, Borel JC, Levy P, Pépin JL. Arterial stiffness in COPD. Chest. 2014;145(4):861–75.
62. Ball MK, Waypa GB, Mungai PT, Nielsen JM, Czech L, Dudley VJ, Beussink L, Dettman RW, Berkelhamer SK, Steinhorn RH, et al. Regulation of Hypoxia-induced pulmonary hypertension by vascular smooth muscle hypoxia-inducible Factor-1α. Am J Respir Crit Care Med. 2014;189(3):314–24.
63. Watz H, Waschki B, Meyer T, Kretschmar G, Kirsten A, Claussen M, Magnussen H. Decreasing cardiac chamber sizes and associated heart dysfunction in COPD: role of hyperinflation. Chest. 2010;138(1):32–8.
64. Barr RG, Bluemke DA, Ahmed FS, Carr JJ, Enright PL, Hoffman EA, Jiang R, Kawut SM, Kronmal RA, Lima JA, et al. Percent emphysema, airflow obstruction, and impaired left ventricular filling. N Engl J Med. 2010;362(3):217–27.
65. Qiu H, Gabrielsen A, Agardh HE, Wan M, Wetterholm A, Wong CH, Hedin U, Swedenborg J, Hansson GK, Samuelsson B, et al. Expression of 5-lipoxygenase and leukotriene A4 hydrolase in human atherosclerotic lesions correlates with symptoms of plaque instability. Proc Natl Acad Sci USA. 2006;103(21):8161–6.
66. Tattersall MC, Guo MY, Korcarz CE, Gepner AD, Kaufman JD, Liu KJ, Barr RG, Donohue KM, McClelland RL, Delaney JA, et al. Asthma predicts Cardiovascular Disease events the multi-ethnic study of atherosclerosis. Arterioscler Thromb Vasc Biol. 2015;35(6):1520–5.
67. Xu MZ, Xu JL, Yang XJ. Asthma and risk of cardiovascular disease or all-cause mortality: a meta-analysis. Ann Saudi Med. 2017;37(2):99–105.
68. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease? Int J Epidemiol. 2004;33(4):743–8.
69. Müllerova H, Agusti A, Erqou S, Mapel DW. Cardiovascular Comorbidity in COPD systematic literature review. Chest. 2013;144(4):1163–78.

70. Rabe KF, Hurst JR, Suissa S. Cardiovascular disease and COPD: dangerous liaisons? Eur Respir Rev. 2018;27(149):32.
71. Ekroos K, Lavrynenko O, Titz B, Pater C, Hoeng J, Ivanov NV. Lipid-based biomarkers for CVD, COPD, and aging - A translational perspective. Prog Lipid Res. 2020;78:12.
72. Cataluña JJ, García MA. [Cardiovascular comorbidity in COPD]. Arch Bronconeumol. 2009;45(Suppl 4):18–23.
73. Mahishale V, Angadi N, Metgudmath V, Eti A, Lolly M, Khan S. Prevalence and impact of diabetes, hypertension, and cardiovascular diseases in chronic obstructive pulmonary diseases: a hospital-based cross-section study. J Translational Intern Med. 2015;3(4):155–60.
74. Magnani JW, Ning HY, Wilkins JT, Lloyd-Jones DM, Allen NB. Educational Attainment and Lifetime Risk of Cardiovascular Disease. JAMA Cardiol. 2024;9(1):45–54.
75. Hamad R, Nguyen TT, Bhattacharya J, Glymour MM, Rehkopf DH. Educational attainment and cardiovascular disease in the United States: a quasi-experimental instrumental variables analysis. PLos Med. 2019;16(6):19.
76. Dégano IR, Marrugat J, Grau M, Salvador-González B, Ramos R, Zamora A, Martí R, Elosua R. The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index. Sci Rep. 2017;7:8.
77. Tayal U, Pompei G, Wilkinson I, Adamson D, Sinha A, Hildick-Smith D, Cubbon R, Garbi M, Ingram TE, Colebourn CL et al. Advancing the access to cardiovascular diagnosis and treatment among women with cardiovascular disease: a joint British Cardiovascular Societies' consensus document. Heart 2024.
78. Yousuf AM, Abdikarim H, Hussein MA, Abdi AN, Warsame HI, Muse AH. Cardiovascular disease prevalence and associated factors in a low-resource setting: a multilevel analysis from Somalia's first demographic health survey. Curr Probl Cardiol. 2024;49(12):12.
79. Sarraju A, Ward A, Chung S, Li J, Scheinker D, Rodríguez F. Machine learning approaches improve risk stratification for secondary cardiovascular disease prevention in multiethnic patients. Open Heart 2021, 8(2).
80. Pollard JD, Haq KT, Lutz KJ, Rogovoy NM, Paternostro KA, Soliman EZ, Maher J, Lima JAC, Musani SK, Tereshchenko LG. Electrocardiogram machine learning for detection of cardiovascular disease in African americans: the Jackson Heart Study. Eur Heart J Digit Health. 2021;2(1):137–51.
81. Sang H, Lee H, Lee M, Park J, Kim S, Woo HG, Rahmati M, Koyanagi A, Smith L, Lee S, et al. Prediction model for cardiovascular disease in patients with diabetes using machine learning derived and validated in two independent Korean cohorts. Sci Rep. 2024;14(1):11.
82. Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, Zhang HJ, Kaplin S, Narasimhan B, Kitai T, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. 2020;10(1):11.

Publisher's note