ESC European Society of Cardiology

European Heart Journal - Digital Health (2024) **00**, 1–13
https://doi.org/10.1093/ehjdh/ztae001

ORIGINAL ARTICLE

Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome

Mitchel A. Molenaar[1,2], Berto J. Bouma [1,2], Folkert W. Asselbergs[1,3], Niels J. Verouden[1,2], Jasper L. Selder[1,2], Steven A.J. Chamuleau[1,2], and Mark J. Schuuring [1,4,5,*]

[1]Department of Cardiology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [2]Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centers, Amsterdam, The Netherlands; [3]Health Data Research UK and Institute of Health Informatics, University College London, London, UK; [4]Department of Cardiology, Circulatory Health Laboratory, University Medical Center Utrecht, Utrecht, The Netherlands; and [5]Circulatory Health UMC Utrecht, Utrecht, The Netherlands

Received 14 August 2023; revised 21 December 2023; accepted 29 December 2023; online publish-ahead-of-print 22 January 2024

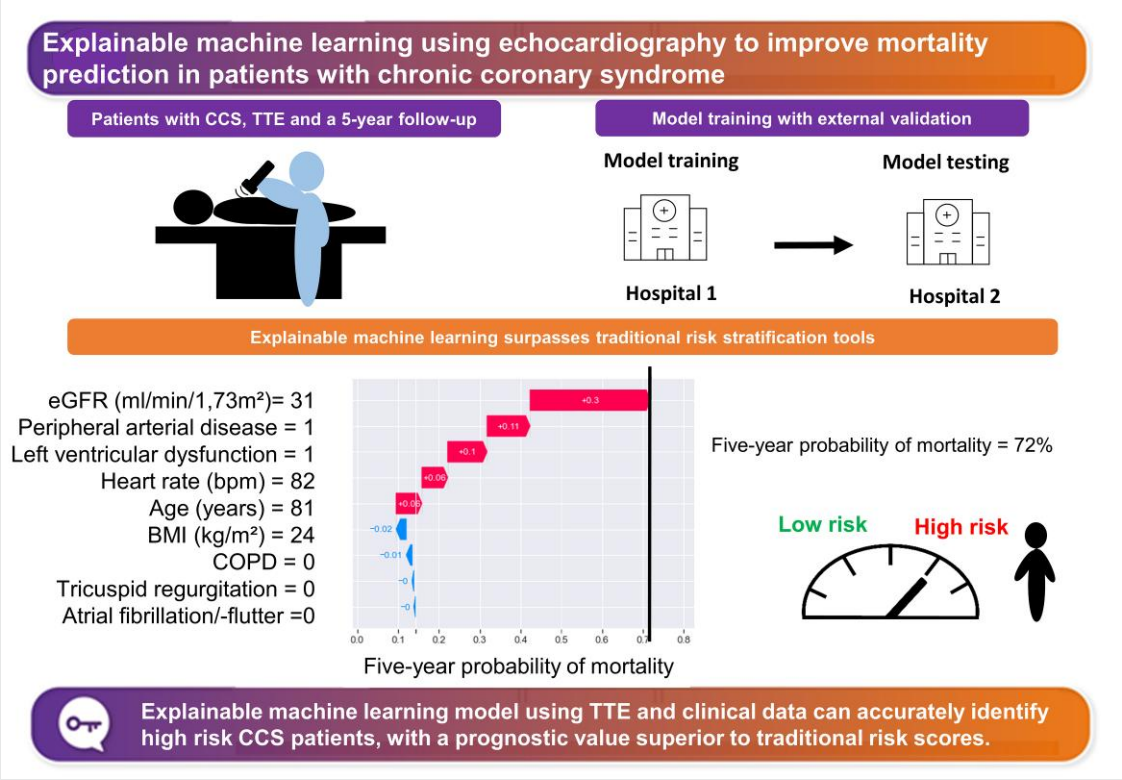
Aims The European Society of Cardiology guidelines recommend risk stratification with limited clinical parameters such as left ventricular (LV) function in patients with chronic coronary syndrome (CCS). Machine learning (ML) methods enable an analysis of complex datasets including transthoracic echocardiography (TTE) studies. We aimed to evaluate the accuracy of ML using clinical and TTE data to predict all-cause 5-year mortality in patients with CCS and to compare its performance with traditional risk stratification scores.

Methods and results Data of consecutive patients with CCS were retrospectively collected if they attended the outpatient clinic of Amsterdam UMC location AMC between 2015 and 2017 and had a TTE assessment of the LV function. An eXtreme Gradient Boosting (XGBoost) model was trained to predict all-cause 5-year mortality. The performance of this ML model was evaluated using data from the Amsterdam UMC location VUmc and compared with the reference standard of traditional risk scores. A total of 1253 patients (775 training set and 478 testing set) were included, of which 176 patients (105 training set and 71 testing set) died during the 5-year follow-up period. The ML model demonstrated a superior performance [area under the receiver operating characteristic curve (AUC) 0.79] compared with traditional risk stratification tools (AUC 0.62–0.76) and showed good external performance. The most important TTE risk predictors included in the ML model were LV dysfunction and significant tricuspid regurgitation.

Conclusion This study demonstrates that an explainable ML model using TTE and clinical data can accurately identify high-risk CCS patients, with a prognostic value superior to traditional risk scores.

* Corresponding author. Tel: +0031887555555, Email: m.j.schuuring-19@umcutrecht.nl

Graphical Abstract



CCS, chronic coronary syndrome; eGFR, estimated glomerular filtration rate; LV, left ventricular; TTE, transthoracic echocardiography.

Keywords Coronary artery disease • Machine learning • Artificial intelligence • Prognosis • Risk • Mortality

Introduction

Chronic coronary syndrome (CCS) is a common cardiovascular condition that affects millions of patients worldwide.[1] Despite receiving medical and interventional treatment, CCS patients have a high rate of cardiovascular events leading to myocardial infarction or mortality in 8% of patients within 5 years.[2] To evaluate the risk of cardiovascular events, the European Society of Cardiology (ESC) guidelines recommend a transthoracic echocardiographic (TTE) assessment of the left ventricular (LV) function in all patients with CCS.[1] Transthoracic echocardiography is the most performed non-invasive cardiac procedure and has unique characteristics such as high temporal resolution, absence of ionizing radiation, portability, and low costs.[3] Furthermore, LV dysfunction has been established as one of the strongest predictors of mortality.[1,4] However, LV function as a stand-alone risk stratifier may not account for all potential risk factors and the complex interactions among them.

Artificial intelligence is a rapidly emerging field and refers to the broad concept of computer systems performing tasks that previously required human intelligence. Machine learning (ML) is a subfield of artificial intelligence in which an algorithm is trained on sample data to perform a specific task (e.g. classification, regression), in order to perform this task on new data.[5,6] Machine learning has shown superior results to predict mortality in patients with CCS using data from coronary computed tomographic angiography (CCTA) or stress cardiac magnetic resonance (CMR) compared with traditional methods.[7,8] Despite these promising results, limited ML studies have used TTE data to predict mortality in patients with CCS.[9]

To work towards an ML model applicable to a larger proportion of the CCS patients, we formulated the research question whether ML using TTE and clinical data can improve risk stratification of patients with CCS. The aim of this study was to investigate the accuracy of ML using clinical and TTE data to predict 5-year mortality in patients with CCS and to compare its performance with traditional risk stratification scores.

Methods

Training cohort

The training cohort was used to train the ML model. This cohort consisted of patients aged 18 years or older diagnosed with CCS who had both an outpatient visit and TTE at the Amsterdam University Medical Center (AUMC), location AMC, the Netherlands, between 2014 and 2017. Chronic coronary syndrome was defined as a clinical presentation of coronary artery disease at the outpatient visit, with the exception of patients in which acute coronary syndrome was the primary clinical presentation.[10] The diagnosis was determined by the treating physician based on history taking, and patients were treated with medication accordingly, in accordance with the ESC guidelines.[10] Additional testing was performed at the discretion of the physician, commonly in patients where medical treatment was ineffective or when there was a need to confirm or refute the CCS diagnosis. Patients were consecutively selected from electronic health records, and their data were extracted from pseudonymized electronic health records and echocardiography reports, which was further described in the study of Molenaar *et al.*[4]

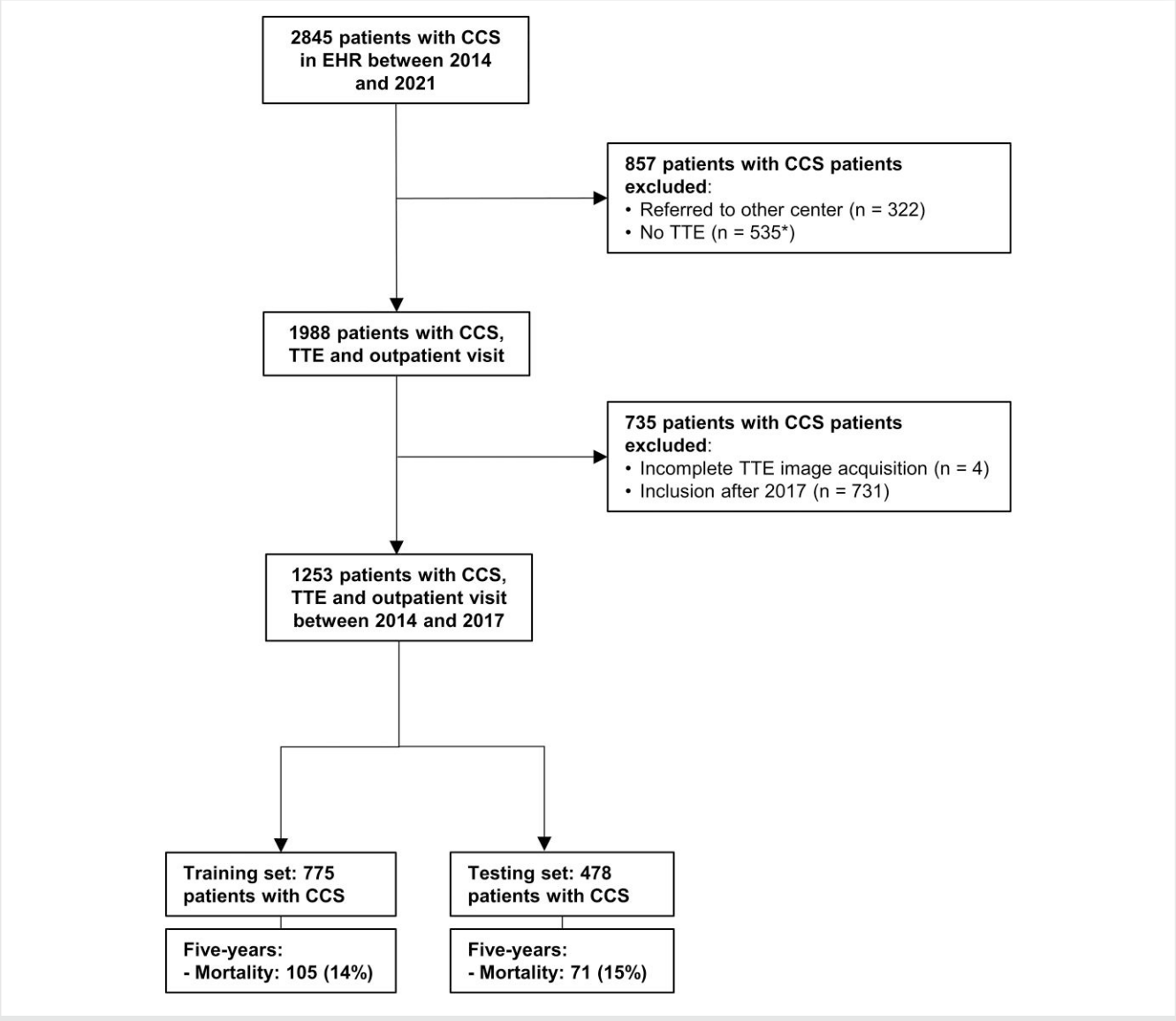


Figure 1 Inclusion flowchart. EHR, electronic health records; CCS, chronic coronary syndrome; TTE, transthoracic echocardiography. *The clinical characteristics of patients without transthoracic echocardiography ($n = 535$) and patients with complete transthoracic echocardiography ($n = 1984$) were comparable, as shown in Supplementary material online, *Table S2*.

Testing cohort

The generalizability of the ML model was assessed by testing it on a cohort of CCS patients in an external centre. This cohort was drawn from a registry of CCS patients between 2014 and 2017 of the AUMC, location VUMC, the Netherlands. The included patients in the testing cohort fulfilled the same inclusion criteria as the training cohort. The local human ethical review board approved the establishment of both registries for study purposes, without the need for written consent.

Outcome

The clinical endpoint was 5-year all-cause mortality, extracted from electronic health records of both medical centres.

Echocardiography data

Chronic coronary syndrome patients underwent two-dimensional TTE with tissue Doppler imaging using various machines, including Vivid 9 (GE Vingmed Ultrasound AS, Horten, Norway) in the training cohort and Philips Epiq, Philips Affiniti, and Philips IE33 (Philips Medical Systems, Best, The Netherlands) in the testing cohort. These TTE assessments were performed by clinical technicians, who followed the recommendations of the ESC guidelines,[11,12] European Association of Cardiovascular Imaging,[13] and standard operating procedure.[14] The TTE images were analysed with vendor-specific software including GE EchoPAC (GE Vingmed Ultrasound AS) in the training cohort and Xcelera (Philips Medical Systems) and TomTec 2D Cardiac Performance Analysis (Munich, Germany) in the testing cohort.

The initial TTE assessment was performed qualitatively by a clinical technician or cardiology resident in routine clinical practice, not performed specifically for this study. Semi-quantitative and quantitative measurements of atrial and ventricular dimensions, right ventricular (RV) function, and valve stenosis or regurgitation were obtained if indicated, especially if clinical decisions were based on these findings.[11–13] In apical two-chamber and four-chamber images, tracings of the LV endocardial borders were performed. The Simpson's biplane method was used to estimate the end-diastolic

Table 1 Baseline characteristics of patients in the training set

Characteristics	All patients (*N* = 775)	Alive (*N* = 670)	Dead (*N* = 105)	*P*-value
Follow-up (years), median (Q1, Q3)	6.46 (5.48, 6.89)	6.55 (5.89, 6.97)	2.27 (1.15, 3.44)	<0.001
Age (years), median (Q1, Q3)	66.00 (58.00, 73.00)	65.00 (57.00, 72.00)	71.00 (64.00, 78.00)	<0.001
Male, *n* (%)	445 (57.4)	383 (57.2)	62 (59.0)	0.797
Risk factors				
Hypertension, *n* (%)	462 (59.6)	397 (59.3)	65 (61.9)	0.683
Diabetes, *n* (%)	232 (29.9)	190 (28.4)	42 (40.0)	0.021
Dyslipidaemia, *n* (%)	260 (33.5)	231 (34.5)	29 (27.6)	0.203
Current or former smoker, *n* (%)	280 (36.1)	235 (35.1)	45 (42.9)	0.151
Family history of CAD, *n* (%)	257 (33.2)	233 (34.8)	24 (22.9)	0.021
Medical history				
Myocardial infarction, *n* (%)	206 (26.6)	166 (24.8)	40 (38.1)	0.006
Recent ACS event[a], *n* (%)	11 (1.4)	9 (1.3)	2 (2.0)	0.326
PCI, *n* (%)	244 (31.5)	206 (30.7)	38 (36.2)	0.315
CABG, *n* (%)	75 (9.7)	56 (8.4)	19 (18.1)	0.003
Valvular repair or replacement, *n* (%)	27 (3.5)	19 (2.8)	8 (7.6)	0.028
Atrial fibrillation/flutter, *n* (%)	88 (11.4)	64 (9.6)	24 (22.9)	<0.001
Stroke, *n* (%)	40 (5.2)	29 (4.3)	11 (10.5)	0.016
COPD, *n* (%)	58 (7.5)	34 (5.1)	24 (22.9)	<0.001
Peripheral arterial disease, *n* (%)	40 (5.2)	26 (3.9)	14 (13.3)	<0.001
Clinical examination				
Chest pain, *n* (%)	427 (55.1)	382 (57.0)	45 (42.9)	0.009
Dyspnoea, *n* (%)	203 (26.2)	163 (24.3)	40 (38.1)	0.004
Other cardiac symptoms, *n* (%)	156 (20.1)	138 (20.6)	18 (17.1)	0.490
Systolic blood pressure (mmHg), mean (SD)	141.02 (21.79)	141.16 (21.26)	140.17 (24.83)	0.687
Heart rate (b.p.m.), mean (SD)	71.74 (13.81)	70.97 (13.46)	76.34 (15.00)	0.001
BMI (kg/m^2), mean (SD)	27.89 (5.64)	28.01 (5.51)	27.12 (6.33)	0.163
Laboratory parameters				
eGFR (mL/min/1.73 m^2)[b], mean (SD)	69.34 (21.00)	71.41 (19.68)	56.32 (24.30)	<0.001
Total cholesterol (mmol/L), mean (SD)	4.57 (1.22)	4.59 (1.22)	4.43 (1.24)	0.262
HDL-cholestrol (mmol/L), mean (SD)	1.32 (0.44)	1.31 (0.42)	1.37 (0.56)	0.271
LDL-cholestrol (mmol/L), mean (SD)	2.93 (1.14)	2.96 (1.14)	2.72 (1.15)	0.071
Triglyceride (mmol/L), mean (SD)	1.55 (1.13)	1.54 (1.17)	1.56 (0.89)	0.925
Echocardiographic measurement				
Left ventricular function, *n* (%)				<0.001
Normal	614 (79.2)	557 (83.1)	57 (54.3)	
Mildly impaired	109 (14.1)	81 (12.1)	28 (26.7)	
Moderately impaired	46 (5.9)	30 (4.5)	16 (15.2)	
Severely impaired	6 (0.8)	2 (0.3)	4 (3.8)	
Right ventricular dysfunction, *n* (%)	52 (6.7)	34 (5.1)	18 (17.1)	<0.001
Right ventricular enlargement, *n* (%)	28 (3.6)	19 (2.8)	9 (8.6)	0.008
Left ventricular enlargement, *n* (%)	90 (11.6)	73 (10.9)	17 (16.2)	0.158
Left atrial enlargement, *n* (%)	338 (43.6)	281 (41.9)	57 (54.3)	0.023
Right atrial enlargement, *n* (%)	91 (11.7)	70 (10.4)	21 (20.0)	0.008
Moderate or severe aortic stenosis, *n* (%)	33 (4.3)	25 (3.7)	8 (7.6)	0.115
Moderate or severe aortic regurgitation, *n* (%)	19 (2.5)	14 (2.1)	5 (4.8)	0.191
Moderate or severe mitral regurgitation, *n* (%)	80 (10.3)	60 (9.0)	20 (19.0)	0.003
Moderate or severe mitral stenosis, *n* (%)	2 (0.3)	2 (0.3)	0 (0.0)	1.000
Moderate or severe pulmonary regurgitation, *n* (%)	4 (0.5)	2 (0.3)	2 (1.9)	0.161
Moderate or severe pulmonary stenosis, *n* (%)	1 (0.1)	1 (0.1)	0 (0.0)	1.000
Moderate or severe tricuspid regurgitation, *n* (%)	56 (7.2)	37 (5.5)	19 (18.1)	<0.001

Continued

Table 1 *Continued*

Characteristics	All patients (*N* = 775)	Alive (*N* = 670)	Dead (*N* = 105)	*P*-value
Diagnostic modalities				
Exercise ECG, *n* (%)	133 (17.2)	123 (18.4)	10 (9.5)	0.036
Stress CMR, *n* (%)	61 (7.9)	54 (8.1)	7 (6.7)	0.766
Myocardial perfusion scan, *n* (%)	298 (38.5)	259 (38.7)	39 (37.1)	0.850
CT coronary angiography, *n* (%)	176 (22.7)	147 (21.9)	29 (27.6)	0.244
Invasive coronary angiography, *n* (%)	463 (59.7)	394 (58.8)	69 (65.7)	0.217
Baseline medication				
Antiplatelet therapy, *n* (%)	516 (66.6)	442 (66.0)	74 (70.5)	0.424
Anticoagulants, *n* (%)	102 (13.2)	77 (11.5)	25 (23.8)	0.001
ACE-inhibitor/ARB, *n* (%)	387 (49.9)	327 (48.8)	60 (57.1)	0.138
Beta-blockers, *n* (%)	468 (60.4)	398 (59.4)	70 (66.7)	0.191
Nitrates or other antianginal drugs, *n* (%)	211 (27.2)	178 (26.6)	33 (31.4)	0.356
Calcium antagonists, *n* (%)	249 (32.1)	213 (31.8)	36 (34.3)	0.692
Diuretics, *n* (%)	219 (28.3)	172 (25.7)	47 (44.8)	<0.001
Statins, *n* (%)	490 (63.2)	422 (63.0)	68 (64.8)	0.809
Insulin, *n* (%)	95 (12.3)	74 (11.0)	21 (20.0)	0.015
Other oral diabetic drugs, *n* (%)	179 (23.1)	151 (22.5)	28 (26.7)	0.419

ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blocker; BMI, body mass index; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; CT, computed tomography; ECG, electrocardiography; eGFR, estimated glomerular filtration rate; PCI, percutaneous coronary intervention; VHD, valvular heart disease.
[a]Within 3 months preceding the presentation at the outpatient clinic.
[b]Calculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation.

and end-systolic LV volumes, through which LV ejection fraction (LVEF) was calculated. Atrial and ventricular enlargements were defined as an increase in the size of their respective chambers. Specifically, left atrial enlargement was defined as a left atrial volume index >34 mL/m^2. Right atrial enlargement was defined as a right atrial volume index ≥30 mL/m^2 for male patients and ≥28 mL/m^2 for female patients. Left ventricular enlargement was defined as a LV end-diastolic dimension >58.4 mm for male patients and >52.2 mm for female patients. Right ventricular enlargement was defined as a RV basal diameter ≥42 mm.[13,15] The results were documented in a TTE report,[13] which was supervised by a dedicated imaging cardiologist who examined the TTE images and made corrections to the TTE report if needed.

The TTE reports were de-identified, and one report was selected for each patient that was closest to the date of the outpatient visit. The following data were extracted from the TTE reports and recorded in the registry: left and right atrial and ventricular enlargements, LV and RV functions, and severity of aortic stenosis, aortic regurgitation, mitral stenosis, mitral regurgitation, tricuspid regurgitation, pulmonary stenosis, and pulmonary regurgitation. Left ventricular dysfunction was defined as mildly to severely impaired LV function (LVEF ≤51% for male and ≤53% for female). Moderately and severely impaired LV function was defined as an LVEF of 30–41% and <30%, respectively.[16] The estimated glomerular filtration rate (eGFR) was calculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation.[17]

Available data and standard risk scores

The following data were available for the training and testing cohorts and were used for model training: demographic data, cardiovascular risk factors, medical history, clinical examination, laboratory measurements, and echocardiographic data. A total of 43 features were used for training, of which 14 features were echocardiographic features. The Framingham risk score[18] and ESC Systematic Coronary Risk Evaluation [SCORE2 (<70 years)/SCORE2-OP (≥70 years)] risk[19,20] were calculated to estimate the 10-year cardiovascular risk for each patient. The variables included in these risk scores are age, gender (only in Framingham risk score), geographical region (only in SCORE2/SCORE2-OP), systolic blood pressure, diabetes (only in SCORE2/SCORE2/OP), smoking status, blood pressure treatment (only in Framingham risk score), and total-cholesterol and HDL-cholesterol.

Feature imputation and selection

Missing values in the training and testing cohorts were imputed by multiple imputation by chained equation with a linear regression model in 10 iterations. This process was repeated for the data of both the training and the testing cohorts separately. An eXtreme Gradient Boosting (XGBoost) model was trained to predict 5-year mortality. The eXtreme Gradient Boosting is a non-linear model that employs an ensemble of decision tree models, which has shown good performance in diverse classification problems.[21,22] A grid search with five times five-fold cross-validation was conducted on the data of the training cohort to tune the hyperparameters of XGBoost (see Supplementary material online, *Table S1*). To obtain the optimal set of features and minimize the risk of overfitting, the tuned XGBoost model was trained using a 10 times 10-fold cross-validation strategy. This involved multiple training cycles in which the model was trained on 90% of the training data and validated on the remaining 10%. This process was repeated 10 times with other randomization of the data to obtain a reliable performance estimate. The feature that most frequently exhibited the lowest importance in the 10 iterations of cross-validation, as observed by the features' importance function of XGBoost, was excluded from the subsequent rounds of model training. To reduce the risk of overfitting and enhance the model's generalizability, the minimal set of features was selected at the point where the model's performance began to decline.

Model training and testing

The ML model was trained with the selected features using 10-fold cross-validation, with the area under the receiver operating characteristic curve (AUC) as the optimization metric. The model was recalibrated using the training data to further improve the calibration (agreement between the observed and the predicted risk of mortality) and was subsequently tested on the testing set. In the testing set, the model was evaluated a thousand times on randomly selected bootstrap samples to obtain a reliable estimate

Table 2 Baseline characteristics of patients in the training and testing cohorts

Characteristics	Training set (*N* = 775)	Testing set (*N* = 478)	*P*-value
Follow-up (years), median (Q1, Q3)	6.46 (5.48, 6.89)	6.34 (5.35, 6.92)	0.403
Age (years), median (Q1, Q3)	66.00 (58.00, 73.00)	66.00 (58.00, 75.00)	0.208
Male, *n* (%)	445 (57.4)	290 (60.7)	0.282
Risk factors			
Hypertension, *n* (%)	462 (59.6)	246 (51.5)	0.006
Diabetes, *n* (%)	232 (29.9)	122 (25.5)	0.105
Dyslipidaemia, *n* (%)	260 (33.5)	127 (26.6)	0.011
Current or former smoker, *n* (%)	280 (36.1)	179 (37.4)	0.682
Family history of CAD, *n* (%)	257 (33.2)	155 (32.4)	0.836
Medical history			
Myocardial infarction, *n* (%)	206 (26.6)	134 (28.0)	0.620
Recent ACS event[a], *n* (%)	11 (1.4)	6 (1.2)	0.404
PCI, *n* (%)	244 (31.5)	190 (39.7)	0.003
CABG, *n* (%)	75 (9.7)	68 (14.2)	0.018
Valvular repair or replacement, *n* (%)	27 (3.5)	14 (2.9)	0.709
Atrial fibrillation/flutter, *n* (%)	88 (11.4)	53 (11.1)	0.958
Stroke, *n* (%)	40 (5.2)	25 (5.2)	1.000
COPD, *n* (%)	58 (7.5)	32 (6.7)	0.680
Peripheral arterial disease, *n* (%)	40 (5.2)	17 (3.6)	0.236
Clinical examination			
Chest pain, *n* (%)	427 (55.1)	293 (61.3)	0.036
Dyspnoea, *n* (%)	203 (26.2)	169 (35.4)	0.001
Other cardiac symptoms, *n* (%)	156 (20.1)	101 (21.1)	0.723
Systolic blood pressure (mmHg), mean (SD)	141.02 (21.79)	137.66 (22.67)	0.029
Heart rate (b.p.m.), mean (SD)	71.74 (13.81)	70.27 (13.92)	0.130
BMI (kg/m^2), mean (SD)	27.89 (5.64)	27.48 (4.62)	0.294
Laboratory parameters			
eGFR (mL/min/1.73 m^2)[b], mean (SD)	69.34 (21.00)	67.46 (20.83)	0.130
Total cholesterol (mmol/L), mean (SD)	4.57 (1.22)	4.50 (1.22)	0.414
HDL-cholestrol (mmol/L), mean (SD)	1.32 (0.44)	1.35 (0.49)	0.317
LDL-cholestrol (mmol/L), mean (SD)	2.93 (1.14)	2.78 (1.11)	0.063
Triglyceride (mmol/L), mean (SD)	1.55 (1.13)	1.69 (0.96)	0.054
Echocardiographic measurement			
Left ventricular function, *n* (%)			<0.001
Normal	614 (79.2)	301 (63.0)	
Mildly impaired	109 (14.1)	105 (22.0)	
Moderately impaired	46 (5.9)	54 (11.3)	
Severely impaired	6 (0.8)	18 (3.8)	
Right ventricular dysfunction, *n* (%)	52 (6.7)	59 (12.3)	0.001
Right ventricular enlargement, *n* (%)	28 (3.6)	18 (3.8)	1.000
Left ventricular enlargement, *n* (%)	90 (11.6)	27 (5.6)	0.001
Left atrial enlargement, *n* (%)	338 (43.6)	111 (23.2)	<0.001
Right atrial enlargement, *n* (%)	91 (11.7)	28 (5.9)	0.001
Moderate or severe aortic stenosis, *n* (%)	33 (4.3)	27 (5.6)	0.325
Moderate or severe aortic regurgitation, *n* (%)	19 (2.5)	17 (3.6)	0.335
Moderate or severe mitral regurgitation, *n* (%)	80 (10.3)	31 (6.5)	0.026
Moderate or severe mitral stenosis, *n* (%)	2 (0.3)	3 (0.6)	0.585
Moderate or severe pulmonary regurgitation, *n* (%)	4 (0.5)	0 (0.0)	0.290
Moderate or severe pulmonary stenosis, *n* (%)	1 (0.1)	0 (0.0)	1.000
Moderate or severe tricuspid regurgitation, *n* (%)	56 (7.2)	31 (6.5)	0.699

Continued

Table 2 *Continued*

Characteristics	Training set (*N* = 775)	Testing set (*N* = 478)	*P*-value
Diagnostic modalities			
Exercise ECG, *n* (%)	133 (17.2)	84 (17.6)	0.912
Stress CMR, *n* (%)	61 (7.9)	7 (1.5)	<0.001
Myocardial perfusion scan, *n* (%)	298 (38.5)	186 (38.9)	0.918
CT coronary angiography, *n* (%)	176 (22.7)	146 (30.5)	0.003
Invasive coronary angiography, *n* (%)	463 (59.7)	338 (70.7)	<0.001
Baseline medication			
Antiplatelet therapy, *n* (%)	516 (66.6)	297 (62.1)	0.123
Anticoagulants, *n* (%)	102 (13.2)	69 (14.4)	0.580
ACE-inhibitor/ARB, *n* (%)	387 (49.9)	227 (47.5)	0.434
Beta-blockers, *n* (%)	468 (60.4)	254 (53.1)	0.014
Nitrates or other antianginal drugs, *n* (%)	211 (27.2)	151 (31.6)	0.112
Calcium antagonists, *n* (%)	249 (32.1)	132 (27.6)	0.104
Diuretics, *n* (%)	219 (28.3)	129 (27.0)	0.672
Statins, *n* (%)	490 (63.2)	293 (61.3)	0.532
Insulin, *n* (%)	95 (12.3)	55 (11.5)	0.758
Other oral diabetic drugs, *n* (%)	179 (23.1)	74 (15.5)	0.001

ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blocker; BMI, body mass index; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; CT, computed tomography; ECG, electrocardiography; eGFR, estimated glomerular filtration rate; PCI, percutaneous coronary intervention; VHD, valvular heart disease.
[a]Within 3 months preceding the presentation at the outpatient clinic.
[b]Calculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation.

Table 3 Follow-up of patients in the training and testing sets

Characteristics	Training set (*N* = 775)	Testing set (*N* = 478)	*P*-value
Revascularization, *n* (%)	199 (25.7)	172 (36.0)	<0.001
Valvular repair or replacement, *n* (%)	18 (2.3)	12 (2.5)	0.983
Mortality, *n* (%)	105 (13.5)	71 (14.9)	0.574

Treatment and mortality during the 5-year follow-up period for patient in the training and testing sets.

of performance. The ML model was trained using Python version 3.8 on a Windows-based computer (CPU: 2.3 GHz, RAM: 8 GB).

The impact of a feature on the ML model's outcome was further illustrated with Shapley Additive exPlanations (SHAP) plots.[23] These plots provide insight into the decision-making process of the model, even for a specific patient, by illustrating both the importance and the impact direction (positive/negative) of a feature compared with a baseline value.

Statistical analysis

Descriptive statistics were expressed as mean values with standard deviation (SD) for normally distributed data and a median with an interquartile range for non-normally distributed data. Nominal or ordinal data were expressed with numbers with percentages. The Shapiro–Wilk test was used to test for normality. The Student's *t*-test or Mann–Whitney *U* test was performed for between-group comparisons of continuous data, as appropriate. For categorical variables either a Pearson's χ^2 test or a Fisher's exact test was performed.

The ML-based risk score was compared with the performance of the following benchmark scores: (i) LV dysfunction; (ii) Framingham risk score;[18] (iii) SCORE2/SCORE2-OP;[19,20] and (iv) a Cox-based risk score, a commonly used model in survival analysis.[24] The AUC values of these scores were reported with corresponding 95% confidence intervals (CIs). The AUC of the ML model was compared with the AUC of the traditional risk scores according to DeLong's test.[25] To derive the Cox-based risk score, a multivariable Cox survival analysis was performed on data in the training set using stepwise backward selection minimizing the Akaike information criterion. With this approach, the most significant predictors were selected in the final Cox model, and the inclusion of non-significant factors was avoided. The natural logarithm of the adjusted hazard ratio for each selected variable served as a coefficient in the Cox-based risk score. To calculate the Cox-based risk for each patient, the coefficients were multiplied by the corresponding data values, and the resulting products were summed. Precision-recall curves were plotted, which provide an insight into the performance of the model, especially when the number of patients in the classes (mortality/non-mortality) are imbalanced.

The calibration of the models was assessed by plotting the predicted risk against the true proportion of mortality across multiple risk categories. In a calibrated model, the predicted proportion of mortality matches the true proportion in each category. The Brier score was calculated as a quantitative measure of the accuracy of the model's predictions. The Brier score ranges between 0 (accurate prediction) and 1 (inaccurate prediction).

The prognostic value of the ML score was further evaluated using survival analysis. A cut-off value was determined for the ML score at which the true-positive rate minus the false-positive rate was maximal. Patients in the testing set with an ML score greater than the cut-off value were classified as high-risk patients, while patients with an ML score equal to or lower than the cut-off value were classified as low-risk patients. Kaplan–Meier curves were plotted to analyse the survival of patients categorized into high-risk and low-risk groups. The prognostic value of the ML score was compared with LV dysfunction by calculating the unadjusted hazard ratios and the log-rank test.

This study meets all CODE-EHR minimum framework standards for the use of healthcare data for clinical research.[26] Statistical analyses were

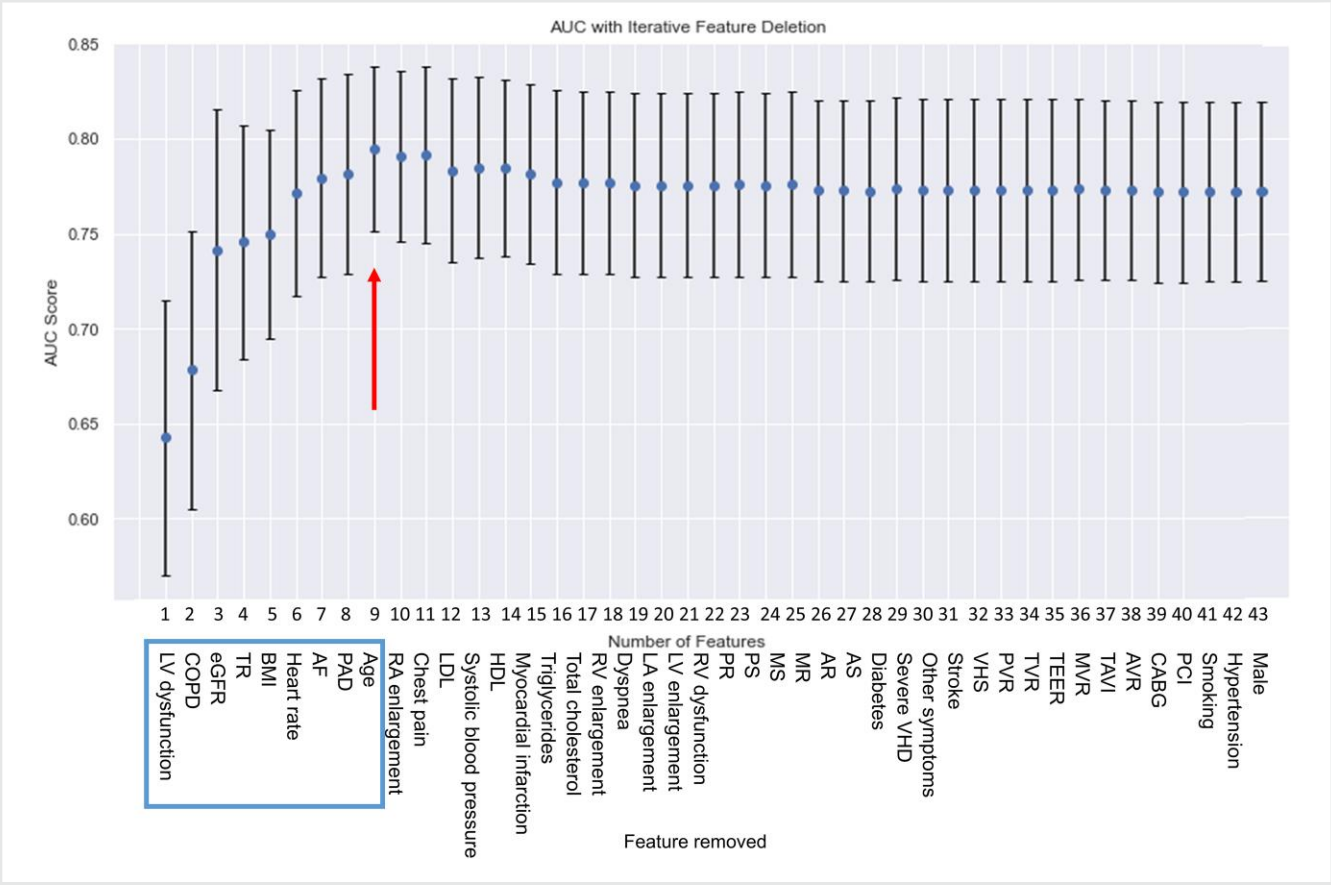


Figure 2 Number of features versus discriminative performance of machine learning model. The first model was trained with all features shown on the *x*-axis using a 10 times 10-fold cross-validation strategy on the training set. The feature that most frequently exhibited the lowest importance in the 10 iterations of cross-validation, as indicated by the features importance function of eXtreme Gradient Boosting, was excluded from the subsequent model training process. The minimal set of features was selected at which the model's discriminative performance began to decline. The performance of the model began to decline when fewer than nine features were included in the model, as indicated by the arrow. Only features surrounded by the box were included in the final model. AF, atrial fibrillation/flutter; AR, moderate or severe aortic regurgitation; AS, moderate or severe aortic stenosis; AUC, area under the receiver operating characteristic curve; AVR, surgical aortic valve replacement; BMI, body mass index; CABG, coronary artery bypass graft; COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate; LA, left atrial; LV, left ventricular; MI, myocardial infarction; MR, moderate or severe mitral regurgitation; MS, moderate or severe mitral stenosis; MVR, surgical mitral valve replacement; PAD, peripheral arterial disease; PCI, percutaneous coronary intervention; PR, moderate or severe pulmonary regurgitation; PS, moderate or severe pulmonary stenosis; PVR, surgical pulmonary valve replacement; RA, right atrial; RV, right ventricular; TAVR, transcatheter aortic valve replacement; TEER, transcatheter edge-to-edge repair; TR, moderate or severe tricuspid regurgitation; TVR, surgical tricuspid valve replacement; VHD, moderate or severe valvular heart disease; VHS, valvular heart surgery.

performed in Python version 3.8 and RStudio version 2022.07.0 (RStudio Team, Boston, MA, USA) using R-version 4.1.3 (R Core Team, Vienna, Austria). A *P*-value <0.05 was considered statistically significant.

Results

Study population

A total of 2845 patients with CCS were screened in both tertiary centres. Among them, 1988 patients had an outpatient visit and underwent TTE between 2014 and 2021. Patients who were excluded due to the absence of a TTE assessment ($n = 535$) had baseline characteristics comparable with patients with a TTE assessment (see Supplementary material online, *Table S2*). The only significant difference was in age, which was higher in the group without a TTE assessment (68 vs. 65 years, $P < 0.001$). After an exclusion of 735 patients who did not have an outpatient visit before 2017 ($n = 731$) or did not have complete TTE acquisition ($n = 4$), 1253 patients were included (775 patients in the training set and 478 patients in the testing set). The flowchart is depicted in *Figure 1*.

Baseline characteristics and follow-up

The baseline characteristics of the patients in the training and testing sets are shown in *Tables 1* and *2*, respectively. Patients in the training set had a median age of 66 years, and 58% of them were male. Additional invasive coronary angiography was most frequently (60%) performed in the diagnostic process, followed by a myocardial perfusion scan (39%) and computed tomography (CT) coronary angiography (23%). Hypertension was the most common risk factor (60%), followed by smoking (36%) and dyslipidaemia (34%). Prior myocardial infarction was reported in 206 patients (27%). Acute coronary syndrome was

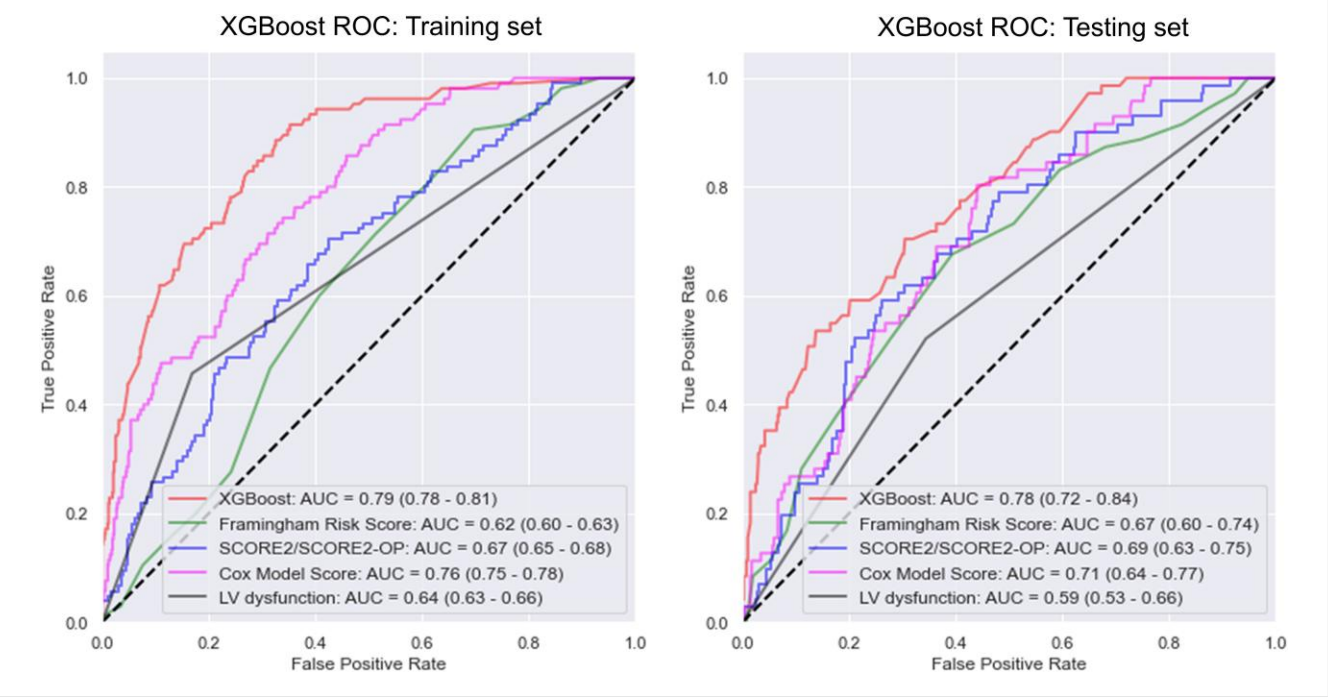


Figure 3 Discriminative performance of the machine learning model and traditional risk scores. The false-positive rate is plotted against the true-positive rate across a range of classification thresholds for the machine learning model (eXtreme Gradient Boosting) and traditional risk scores (training set: left figure, testing set: right figure). The machine learning model exhibited a superior discriminative performance compared with the traditional risk scores for both the training set ($n = 775$, all $P < 0.001$) and the testing set ($n = 478$, all $P < 0.03$). AUC, area under the receiver operating characteristic curve; LV, left ventricular; ROC, receiver operating curve; XGBoost, eXtreme Gradient Boosting.

reported in 11 patients (1%) during the 3 months preceding the presentation at the outpatient clinic. Revascularization was performed by percutaneous coronary intervention in 244 patients (32%) and by coronary artery bypass grafting (CABG) in 75 patients (10%). Patients had a mean body mass index (BMI) of 28 kg/m^2, and chest pain was reported in 427 patients (55%). Left ventricular dysfunction was reported in 161 patients (21%), with severe dysfunction in 6 (1%) patients. During the 5-year follow-up period, 199 patients received revascularization (26%), 18 patients received valvular heart repair or replacement (2%), and 105 (14%) patients died (*Table 3*). A total of 2% of the data were missing, as shown in Supplementary material online, *Table S3*.

Feature selection

In the final ML model, a total of nine features were included. The discriminative performance of the ML model began to decline when fewer than nine features were included in the ML model, as demonstrated in *Figure 2*. The clinical features included in the final ML model were eGFR, age, heart rate, BMI, chronic obstructive pulmonary disease, atrial fibrillation/flutter atrial, and peripheral arterial disease. In addition, the TTE features, LV dysfunction and moderate or severe tricuspid regurgitation, were included in the model. The final ML model was trained using these nine features and subsequently evaluated on the testing set.

Prediction of mortality

The ML model [AUC: 0.79 (95% CI 0.78–0.81)] demonstrated superior discriminative performance compared with LV dysfunction [AUC: 0.64 (95% CI 0.63–0.66)], Framingham risk score [AUC: 0.62 (95% CI 0.60–0.63)], SCORE2/SCORE2-OP [AUC: 0.67 (95% CI 0.65–0.68)], and Cox-based risk score [AUC: 0.76 (95% CI 0.75–0.78), all $P < 0.001$]. The discriminative performance of the models is shown in *Figure 3*. Compared with the traditional risk scores, the precision-recall curve of the ML model showed the best trade-off between precision and recall across different classification thresholds (see Supplementary material online, *Figure S1*). The variables in the final Cox-based model included age, diabetes, current or former smoker, chronic obstructive pulmonary disease, chest pain, renal function (eGFR < 60 mL/min/1.73 m^2), LV dysfunction (mildly to severely impaired function), and atrial fibrillation/flutter.

External validation

The baseline characteristics and follow-up of patients in the testing set are shown in *Tables 2* and *3*, respectively. Patients in the testing set had comparable characteristics with those in the training set in terms of age, gender, and laboratory parameters. In the diagnostic process of patients in the testing set, additional invasive coronary angiography (71 vs. 60%) and CT coronary angiography (31 vs. 23%) were more frequently performed compared with the training set. Stress cardiac magnetic imaging was performed more frequently in the training set (2 vs. 8%). Patients in the testing set had lower rates of hypertension (52 vs. 60%) and dyslipidaemia (27 vs. 34%) compared with the training set, but higher rates of revascularization (percutaneous coronary intervention: 40 vs. 32%, CABG: 14 vs. 10%) and LV dysfunction (37 vs. 21%). Revascularization was more often reported in patients in the testing set during the 5-year follow-up period (36 vs. 26%). Mortality was reported in 71 patients (15%) in the testing set, which was comparable with patients in the training set. A total of 4% of the data were missing, as shown in Supplementary material online, *Table S3*.

In external validation, the highest AUC was observed for the ML model [AUC: 0.78 (95% CI 0.72–0.84)], followed by the Cox-based score [AUC: 0.71 (95% CI 0.64–0.77), $P = 0.002$], SCORE2/SCORE2-OP [AUC: 0.69 (95% CI 0.63–0.75), $P = 0.03$], Framingham

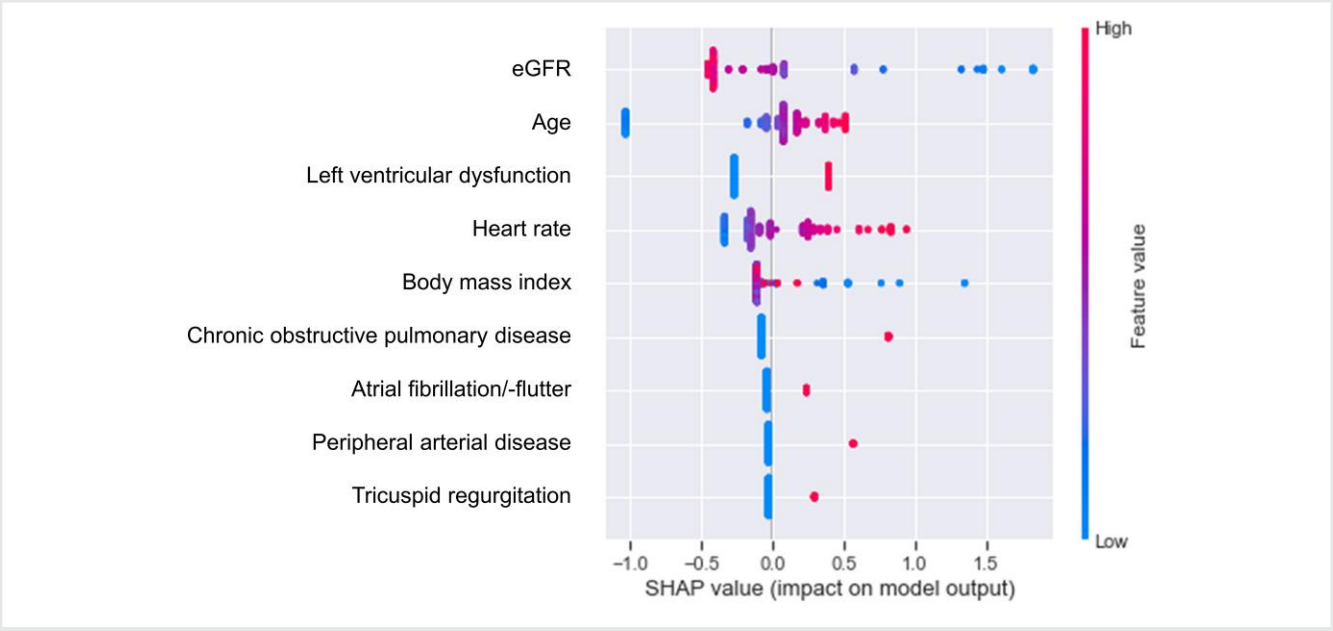


Figure 4 Shapley Additive exPlanations plot of feature importance in the final machine learning model. The importance of features is shown in increasing order on the *y*-axis (estimated glomerular filtration rate is the most important). The relative impact of these features on the model output is depicted on the *x*-axis (the right of 0.0 means increased risk and the left of 0.0 means reduced risk). The value of the feature is shown in colors. For example, a low estimated glomerular filtration rate is associated with a higher risk of mortality (on the right side of the *x*-axis), while a high estimated glomerular filtration rate (on the left side of the *x*-axis) has a protective effect. eGFR, estimated glomerular filtration rate; SHAP, Shapley Additive exPlanations.

risk score [AUC: 0.67 (95% CI 0.60–0.74), $P = 0.006$], and LV dysfunction [AUC: 0.59 (95% CI 0.53–0.66), $P < 0.001$; *Figure 3*).

Individual risk prediction: explainable machine learning

As depicted in *Figure 4*, eGFR, age, and LV dysfunction were the most important features in the final ML model. The individual risk predictions of two patients (high and low risks) in the testing set are shown in *Figure 5*. As illustrated in this figure, the contribution of each feature to the output of the model was different for each patient.

Calibration

The predicted probabilities from the ML model were aligned with the observed mortality rates in each risk category. The Brier scores were 0.08 (95% CI 0.07–0.10) for the training set and 0.10 (95% CI 0.09–0.12) for the testing set, as shown in Supplementary material online, *Figure S2*.

Risk stratification by the machine learning model

A probability cut-off value of 14% was chosen to distinguish high-risk from low-risk patients. At this specific cut-off value, the ML score demonstrated a good prognostic value [unadjusted hazard ratio: 4.7 (95% CI 3.1–7.2)], as shown in *Figure 6*. The ML score exhibited superior discriminative ability compared with LV dysfunction [unadjusted hazard ratio: 1.9 (95% CI 1.3–3.0)] in distinguishing individuals at high and low risks of mortality.

Discussion

This study demonstrates that an explainable ML model using TTE and clinical data can accurately identify CCS patients with a high risk of 5-year mortality. The employed ML model, trained on 775 patients, had a prognostic value superior to LV dysfunction and other traditional risk scores. These findings suggest that ML may support clinicians in assessing the individual risk of mortality of CCS patients.

Machine learning for risk stratification

The prevalence and mortality rates of CCS are exhibiting an upward trend, posing a major challenge for risk stratification of these patients.[27] To allocate healthcare resources to patients with a greater risk of cardiovascular events, accurate risk stratification is essential. However, in our study, risk stratification utilizing the gold standard LV function demonstrated a low predictive power. The incorporation of multiple variables as observed in the Framingham risk score, SCORE2/SCORE2-OP, and Cox model score resulted in more accurate discrimination compared with LV function, but not to the same extent as the ML model. In external validation, the ML model exhibited the highest discriminative performance, which suggests that risk stratification of CCS patients necessitates models that can incorporate non-linear relationships and complex interactions.

To our knowledge, this is the first study that investigated the performance of explainable ML in patients with CCS using TTE and clinical data. The use of TTE data in prediction models offers several advantages since it is widely used, non-invasive, and inexpensive.[3] Previous studies have included data from other imaging modalities in their ML models, which showed performances in line with our findings. Motwani *et al.*[8] trained an ML model to predict 5-year mortality in 10 030 patients with suspected CCS who underwent CCTA.

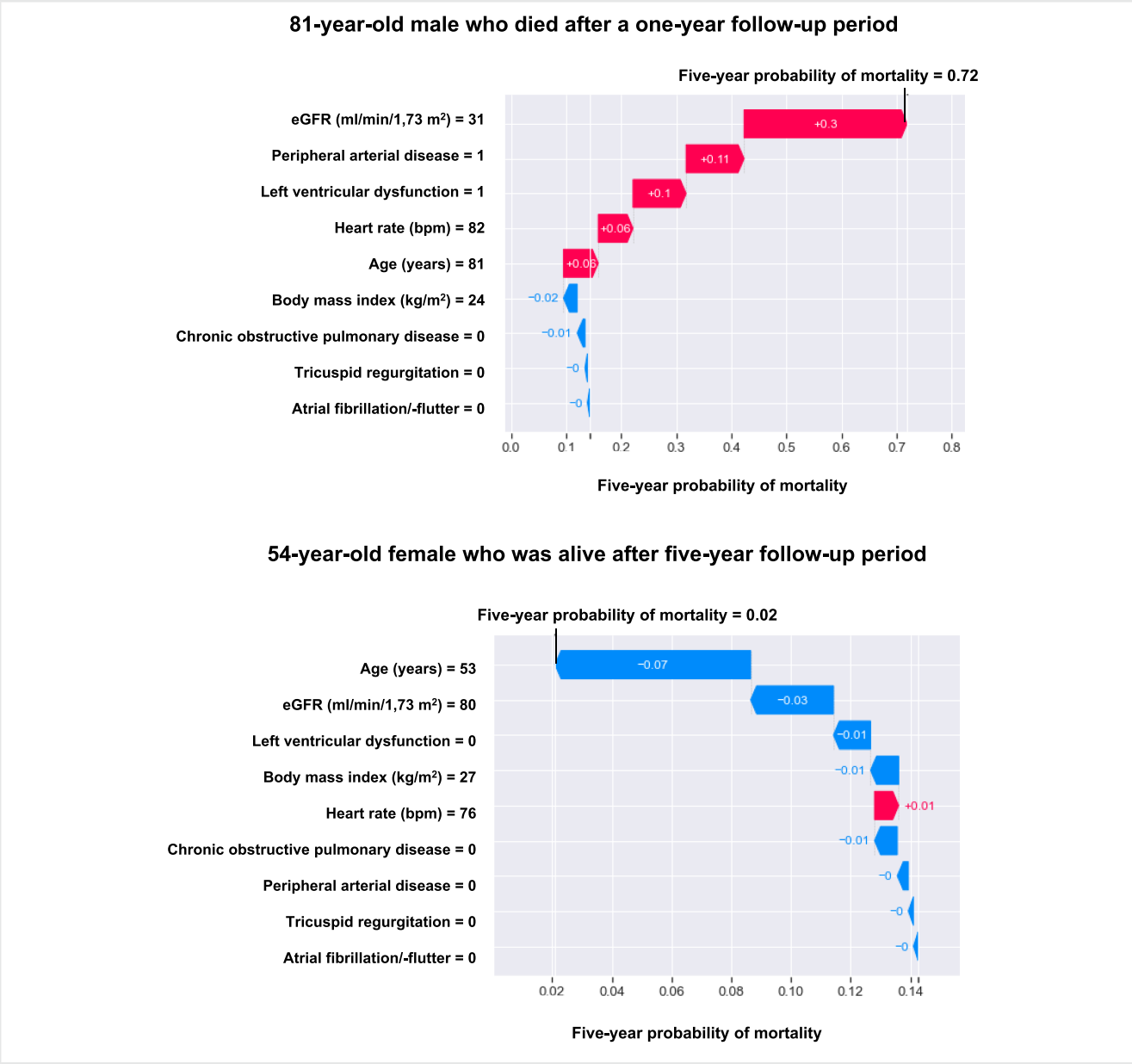


Figure 5 Feature contribution to predict five-year mortality in a machine learning model. Individual predictions are shown for an 81-year-old male patient (top figure) who died within 1 year after presenting at the outpatient clinic and for a 54-year-old female patient (bottom figure) who was alive after the 5-year follow-up period. The impact of the features on the output of the machine learning model is ranked from top (most impact) to bottom (least impact). The size and direction of the arrows indicate how each variable impacts 5-year mortality. An arrow pointing to left indicates a reduction in risk and arrow pointing to the right indicates an increase in risk. The final mortality prediction for the individual patient is determined by the summed impact of all features. eGFR, estimated glomerular filtration rate.

Both clinical and CCTA data were included in the ML model, which exhibited a higher AUC (0.79) compared with the Framingham risk score and CCTA severity scores. More recently, Pezel *et al.*[7] trained an ML model to predict mortality using clinical and stress magnetic resonance (CMR) in 31 752 patients with suspected or known CCS. The authors showed that the ML model was able to predict 10-year mortality more accurately (AUC 0.76) compared with traditional risk scores. These studies emphasize the potential role of ML in addressing the challenge of overseeing the growing number of imaging and clinical variables for risk assessment of the individual patient.[7]

The feature moderate or severe tricuspid regurgitation was included as a predictor in the ML model. In a prior study,[4] we demonstrated the prognostic value of tricuspid regurgitation in patients with CCS, independent of LV dysfunction. These findings emphasize the importance of incorporating echocardiographic features of cardiac structure and function in risk prediction models.

Explainable machine learning

In this study, a tree-based ML model was trained that provides information about the model's decision-making by means of SHAP values.

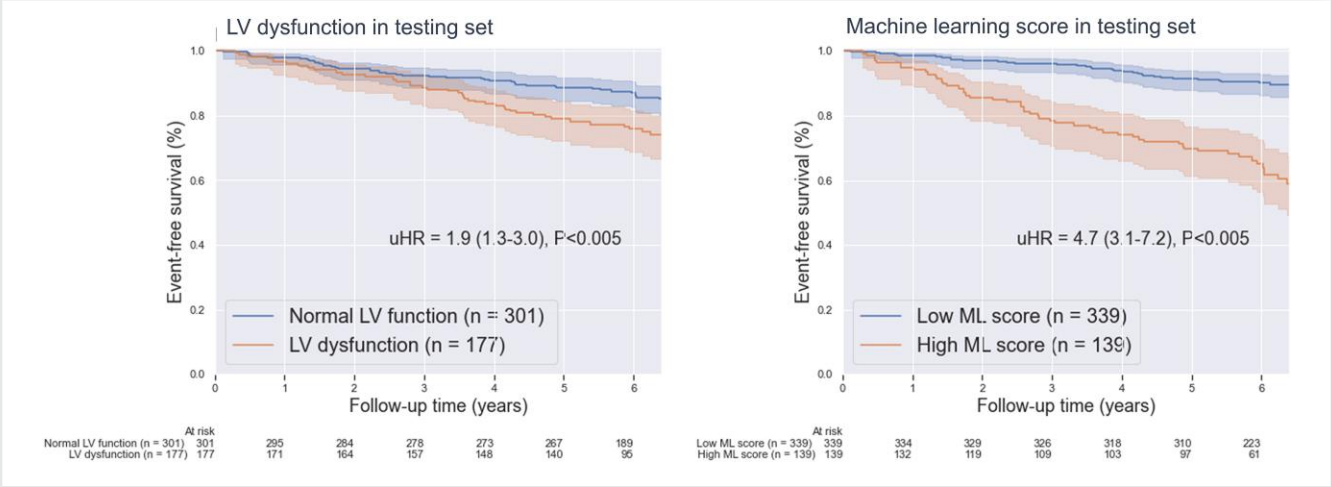


Figure 6 Kaplan–Meier survival curves for the machine learning score and LV dysfunction. The machine learning score (right figure) had a superior discriminative ability compared with left ventricular dysfunction (left figure) in distinguishing individuals at high and low risks of mortality. This score was calculated for patients in the testing set ($n = 478$). LV, left ventricular; ML, machine learning; uHR, unadjusted hazard ratio.

To date, there is an ongoing debate whether currently available explainability techniques are sufficient and a prerequisite to inform clinicians about decisions for the individual patient.[28] Explanations regarding the model's output are only approximations, and the underlying model may be incorrect, which introduces bias in explainability techniques. Despite these acknowledged limitations, it is important to recognize that the explainability technique used in this study is one of the most effective approaches currently available to provide an insight into the decision-making process of the model.[29] To further enhance trustworthiness, further efforts are needed to provide information about the certainty of a decision.

Study limitations

Several remarks can be made about this study. Patients without a TTE assessment at the investigated centres were excluded from analysis, which may have led to an unrepresentative CCS population. However, these patients had comparable baseline characteristics with those with TTE assessment, which reduces the risk of selection bias. The TTE assessments in this study were conducted as part of standard clinical practice. The reproducibility of these TTE assessments was not evaluated. In our study, BMI was chosen as an established risk factor as recommended by the ESC guidelines, despite the growing evidence that waist-to-height ratio may be a more accurate predictor of mortality.[10,30–32]

The developed ML model was trained to predict all-cause 5-year mortality. The cause of mortality could not be obtained for all patients and was therefore not further differentiated. There is currently a lack of traditional 5-year mortality risk scores for patients with CCS. Therefore, we did the same as Motwani *et al.*[8] did, who compared the ML score with traditional risk scores that are designed to estimate the 10-year risk of cardiovascular events. The number of patients included in this study was limited to patients from two tertiary centres and was low compared with that in previous ML studies. The strength of our study was the testing of the ML model in an external centre, which is a prerequisite to determine the generalizability of a prediction model.[33] To prevent the model from overfitting, we iteratively eliminated the least important features until the model's performance began to decline. The final ML model generalized well (AUC testing set 0.78) on unseen data and outperformed the LV function and other traditional risk scores. Despite the lack of quantitative parameters of TTE (e.g. parameters of diastolic function) and the limited number of variables in the ML model, its performance remained consistent in the testing set even when there were minor differences in patient characteristics compared with those in the training set. Datasets with a larger number of patients and the inclusion of quantitative parameters of TTE may facilitate the identification of more complex patterns related to mortality, which may further enhance the risk prediction in patients with CCS.

Our study primarily focused on the potential of ML-based mortality prediction using clinical and TTE data, and a comparison between this ML–TTE score and ML-based mortality prediction models on CCTA or CMR data was not within the scope of our research. A comparative analysis between ML-based TTE prediction and ML-based CCTA and CMR scores could provide insights into their relative performance for mortality prediction. In addition, combining these multimodal data in ML models may further enhance risk prediction. Once such a model is implemented in a clinical setting and prospectively validated; the effectiveness of ML-based risk stratification could be evaluated in randomized controlled trials.[34]

Conclusions

This study demonstrates that an explainable ML model using TTE and clinical data can accurately identify CCS patients with a high risk of 5-year mortality, with a prognostic value superior to those of LV dysfunction and other traditional risk scores. These findings are of clinical relevance since they indicate that ML may support clinicians in assessing the individual risk of mortality in CCS patients. Larger datasets are needed to train and validate an ML model for patients with CCS using clinical and TTE data.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

The authors thank Casper Coerkamp, Hafsah Rahimbaksh, Merijn Hofland, Tom Peper, Paulo Heemskerk, Björn van der Ster, and Matthijs Schakel for their support in this project.

Funding

The study was supported by the internal funding institution.

Conflict of interest: F.A. is supported by UCL Hospitals NIHR Biomedical Research Centre, EU Horizon (AI4HF 101080430 and DataTools4Heart 101057849), and Dutch Research Council (MyDigiTwin 628.011.213).

Data availability

The original contributions presented in the study are included in the article/Supplementary material online. Further inquiries can be directed to the corresponding author. The ML model trained in this study is available online: https://github.com/MitchMolenaar/CCSML.

References

1. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, *et al.* 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;**41**:407–477.
2. Sorbets E, Fox KM, Elbez Y, Danchin N, Dorian P, Ferrari R, *et al.* Long-term outcomes of chronic coronary syndrome worldwide: insights from the international CLARIFY registry. *Eur Heart J* 2020;**41**:347–356.
3. Schuuring MJ, Išgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;**8**:648877.
4. Molenaar MA, Bouma BJ, Coerkamp CF, Man JP, Išgum I, Verouden NJ, *et al.* The impact of valvular heart disease in patients with chronic coronary syndrome. *Front Cardiovasc Med* 2023;**10**:1211322.
5. Gill SK, Karwath A, Uh H-W, Cardoso VR, Gu Z, Barsky A, *et al.* Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *Eur Heart J*2023;**44**:713–725.
6. Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, *et al.* Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease. *Curr Cardiol Rep* 2022;**24**:365–376.
7. Pezel T, Sanguineti F, Garot P, Unterseeh T, Champagne S, Toupin S, *et al.* Machine-learning score using stress CMR for death prediction in patients with suspected or known CAD. *JACC Cardiovasc Imaging* 2022;**15**:1900–1913.
8. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, *et al.* Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J* 2017;**38**:500–507.
9. Mishra RK, Tison GH, Fang Q, Scherzer R, Whooley MA, Schiller NB. Association of machine learning-derived phenogroupings of echocardiographic variables with heart failure in stable coronary artery disease: the heart and soul study. *J Am Soc Echocardiogr* 2020;**33**:322–331.e1.
10. Knuuti J. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes the task force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Russ J Cardiol* 2020;**25**:119–180.
11. Vahanian A, Alfieri O, Andreotti F, Antunes MJ, Barón-Esquivias G, Baumgartner H, *et al.* Guidelines on the management of valvular heart disease (version 2012). *Eur Heart J* 2012;**33**:2451–2496.
12. Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, *et al.* 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J* 2017;**38**:2739–2791.
13. Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, *et al.* Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2017;**18**:1301–1310.
14. Bouma BJ, Riezenbos R, Voogel AJ, Veldhorst MH, Jaarsma W, Hrudova J, *et al.* Appropriate use criteria for echocardiography in the Netherlands. *Neth Heart J* 2017;**25**:330–334.
15. Koster J, Pieper PG. Normaalwaarden. In: Hamer JPM, Pieper PG, eds. Praktische echocardiografie. 3rd ed. Houten: Bohn Stafleu van Loghum; 2015, p389–410.
16. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, *et al.* Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 2015;**28**:1–39.e14.
17. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, *et al.* New equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;**150**:604–612.
18. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;**117**:743–753.
19. SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;**42**:2439–2454.
20. SCORE2-OP Working Group and ESC Cardiovascular Risk Collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *Eur Heart J* 2021;**42**:2455–2467.
21. Al'Aref SJ, Singh G, Choi JW, Xu Z, Maliakal G, van Rosendael AR, *et al.* A boosted ensemble algorithm for determination of plaque stability in high-risk patients on coronary CTA. *JACC Cardiovasc Imaging* 2020;**13**:2162–2173.
22. Wenzl FA, Kraler S, Ambler G, Weston C, Herzog SA, Räber L, *et al.* Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *Lancet* 2022;**400**:744–756.
23. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017, p.4768–4777.
24. Cox DR. Regression models and life-tables. *J R Stat Soc B* 1972;**34**:187–202.
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–845.
26. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, *et al.* CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *Eur Heart J* 2022;**43**:3578–3588.
27. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, *et al.* Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J Am Coll Cardiol* 2020;**76**:2982–3021.
28. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;**3**:e745–e750.
29. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021;**113**:103655.
30. Ashwell M, Mayhew L, Richardson J, Rickayzen B. Waist-to-height ratio is more predictive of years of life lost than body mass index. *PLoS One* 2014;**9**:e103483.
31. Abdi Dezfouli R, Mohammadian Khonsari N, Hosseinpour A, Asadi S, Ejtahed H-S, Qorbani M. Waist to height ratio as a simple tool for predicting mortality: a systematic review and meta-analysis. *Int J Obes* 2023;**47**:1286–1301.
32. Cox BD, Whichelow M. Ratio of waist circumference to height is better predictor of death than body mass index. *BMJ* 1996;**313**:1487.
33. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, *et al.* External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;**56**:826–832.
34. Schuuring M, Man J, Chamuleau S. Inclusive health tracking: unlock the true potential of digital health solutions. *JACC: Advances* 2023;**2**:100545.