ORIGINAL ARTICLE

Preventing unnecessary imaging in patients suspect of coronary artery disease through machine learning of electronic health records

L. Malin Overmars [1], **Bram van Es**[1], **Floor Groepenhoff** [1,2], **Mark C.H. De Groot**[1], **Gerard Pasterkamp**[1], **Hester M. den Ruijter**[2], **Wouter W. van Solinge**[1], **Imo E. Hoefer**[1], **and Saskia Haitjema**[1]*

[1]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht, Heidelberglaan 100 3584 CX, the Netherlands; and [2]Laboratory of Experimental Cardiology, University Medical Center Utrecht, Heidelberglaan 100 3584 CX, Utrecht, the Netherlands

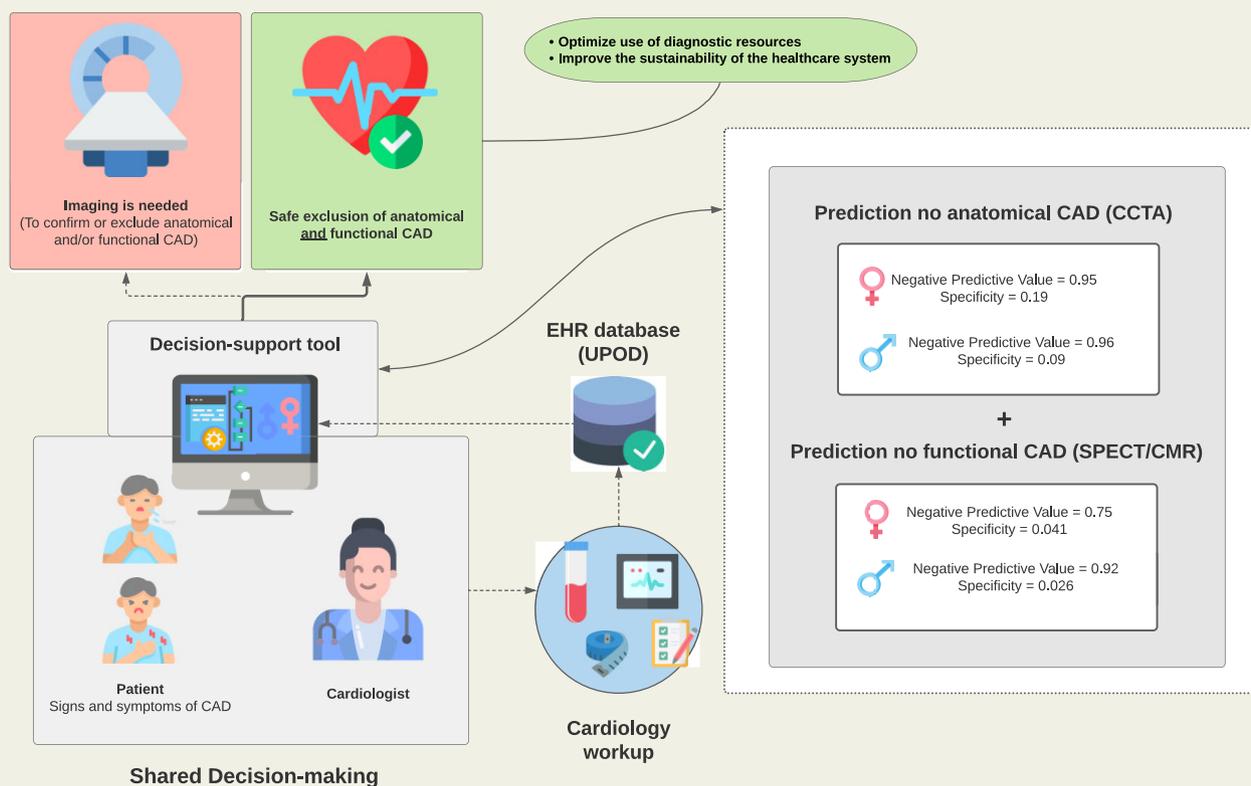|	
Aims	With the ageing European population, the incidence of coronary artery disease (CAD) is expected to rise. This will likely result in an increased imaging use. Symptom recognition can be complicated, as symptoms caused by CAD can be atypical, particularly in women. Early CAD exclusion may help to optimize use of diagnostic resources and thus improve the sustainability of the healthcare system. To develop sex-stratified algorithms, trained on routinely available electronic health records (EHRs), raw electrocardiograms, and haematology data to exclude CAD in patients upfront.
Methods and results	We trained XGBoost algorithms on data from patients from the Utrecht Patient-Oriented Database, who underwent coronary computed tomography angiography (CCTA), and/or stress cardiac magnetic resonance (CMR) imaging, or stress single-photon emission computerized tomography (SPECT) in the UMC Utrecht. Outcomes were extracted from radiology reports. We aimed to maximize negative predictive value (NPV) to minimize the false negative risk with acceptable specificity. Of 6808 CCTA patients (31% female), 1029 females (48%) and 1908 males (45%) had no diagnosis of CAD. Of 3053 CMR/SPECT patients (45% female), 650 females (47%) and 881 males (48%) had no diagnosis of CAD. On the train and test set, the CCTA models achieved NPVs and specificities of 0.95 and 0.19 (females) and 0.96 and 0.09 (males). The CMR/SPECT models achieved NPVs and specificities of 0.75 and 0.041 (females) and 0.92 and 0.026 (males).
Conclusion	Coronary artery disease can be excluded from EHRs with high NPV. Our study demonstrates new possibilities to reduce unnecessary imaging in women and men suspected of CAD.

* Corresponding author. Tel: +31 887550759, Email: s.haitjema@umcutrecht.nl

Graphical Abstract

Imaging is needed
(To confirm or exclude anatomical and/or functional CAD)

Safe exclusion of anatomical **and** functional CAD

• Optimize use of diagnostic resources
• Improve the sustainability of the healthcare system

Prediction no anatomical CAD (CCTA)

♀ Negative Predictive Value = 0.95
Specificity = 0.19

♂ Negative Predictive Value = 0.96
Specificity = 0.09

+

Prediction no functional CAD (SPECT/CMR)

♀ Negative Predictive Value = 0.75
Specificity = 0.041

♂ Negative Predictive Value = 0.92
Specificity = 0.026

Decision-support tool

EHR database (UPOD)

Patient
Signs and symptoms of CAD

Cardiologist

Cardiology workup

Shared Decision-making

Introduction

In 2050, the number of European citizens over 65 years old will be increased by 8% (compared to 2020) to 128 million.[1] With the ageing European population, the incidence of coronary artery disease (CAD) is expected to increase dramatically.[2] Symptom recognition can be complicated, as symptoms caused by CAD are atypical, particularly in women. Hence, the diagnostic workup in patients with atypical symptoms can be time-consuming and expensive and is restricted by the availability of specialized imaging modalities, which are often limited to larger healthcare centres or hospitals.[3–6]

From a diagnostic viewpoint, the main task of the cardiologist is to combine and integrate the results of the diagnostic modalities and conclude whether further cardiologic treatment and care is required or refer to another specialism for further assessment of differential diagnosis and extracardiac causes. To meet the growing healthcare demand and the need to reduce healthcare-related costs, clinical decision support systems (CDSS) to rule out CAD upfront would be of added value. The systematic collection and storage of routine care data in electronic health record (EHR) databases enables the use of

large care datasets for CDSS development.[7] Furthermore, computer systems can integrate larger amounts of data and more variables than human beings, facilitating the exploration of other data sources, such as raw electrocardiogram (ECG) recording and routinely generated haematological biomarkers in CDSS development.[8–11]

Currently, diagnosis of atherosclerosis of the epicardial coronary arteries by coronary computed tomography angiography (CCTA) is relatively straightforward for patients with traditional risk factors and symptoms. However, the diagnosis of non-obstructive (e.g. microvascular disease) yet functionally significant CAD, which is more prevalent in women than in men, is more challenging.[12–15] Patients with chest discomfort and non-obstructive CAD more often show coronary microvascular dysfunction.[16] Previous studies have shown that these patients have a higher risk of major adverse cardiac events and all-cause mortality, and show higher prevalence of stress and psychiatric disturbances, compared to healthy individuals or patients with macrovascular CAD.[14,16–18] Current prediction models have been solely developed to aid the cardiologist in diagnosing or excluding CAD on CCTA, yet did not include the presence of functional CAD, i.e. myocardial ischaemia, in patients without obstructive

CAD.[19–21] Exclusion of both, anatomical CAD (i.e. obstructive) and functional (i.e. non-obstructive) CAD as part of the disease continuum requires a broad and open approach, demanding more comprehensive datasets exceeding conventional risk factors.

To determine the possibility to provide clinical decision support in excluding CAD in patients with chest discomfort, we aimed to develop sex-stratified algorithms using EHRs, including raw ECGs and haematological biomarkers.[22] The resulting algorithms predict the outcome of CAD on CCTA and on stress cardiac magnetic resonance (CMR)/stress single-photon emission computerized tomography (SPECT) and could be implemented in clinical care before imaging is performed.

Methods

Study population and data source

We selected and analysed individual EHR data of patients with available CCTA and/or stress CMR/stress SPECT between February 1997 and September 2019 from the Utrecht Patient-Oriented Database (UPOD). The structure and content of UPOD have been described in more detail elsewhere.[23] In brief, UPOD is an infrastructure of relational databases comprising EHR data for all patients treated at the University Medical Center Utrecht (UMC Utrecht), Utrecht, the Netherlands.

Patients who underwent CCTA, SPECT, and/or CMR were selected based on procedure codes. We excluded: (i) patients under 18 years old, (ii) patients with SPECT and CMR reports without the presence of the terms 'adenosine' or 'regadenoson' to retain only the stress SPECT/CMR scans, (iii) heart transplant patients, (iv) patients who received imaging for other reasons such as ablation procedures, transcatheter aortic valve replacement planning and pulmonary vein mapping by a combination of text mining and manually reviewing radiology reports, and (v) patients with data availability of less than 20%.

For this patient population, we extracted medication prescriptions and dates (ATC codes), laboratory measurements, complete full blood counts, cell size, cell complexity, and cell fluorescence [including research-only parameters from Sapphire Cell-Dyn blood cell analyzers (Abbott Hematology, Santa Clara, CA, USA), raw ECG data (electrical activity of the heart as measured in millivolts on two planes), clinical measurements, and clinical correspondence (radiology reports and cardiology letters)] (Supplementary material online, Table S1).

UPOD data acquisition and management is following current regulations concerning privacy and ethics. All data were pseudonymized before use in the study. The current study was conducted under the declaration of Helsinki.

Feature extraction

The dataset used for model development comprised of patients' most recent data, collected within 12 months before the scan date. Data from and before baseline year were used to define cardiovascular risk factors, including medication prescriptions (diabetes mellitus: insulin or oral hypoglycaemics ATC A10, hypertension: antihypertensives ATC C02, hypercholesterolaemia: lipid modifying agents ATC C10), measurement data (hypertension: >3 measurements of which mean systolic blood pressure > 140 or mean diastolic blood pressure > 90), and laboratory data (hypercholesterolaemia: total cholesterol > 6 and LDL cholesterol > 3.5). We extracted smoking status using a custom-made data mining algorithm[24] (Supplementary material online, Table S1).

Outcome definition

Outcomes were text-mined from radiology reports using a natural language processing procedure based on regular expressions to annotate all affirmative and negative mentions of CAD. The following criteria were used: (i) absence of CAD on CCTA: negative mentions of plaque, stenosis, and/or coronary calcifications, and (ii) absence of functional CAD on SPECT/CMR: a summed stress score <4 or negative mentions of ischaemia and/or perfusion defects. These outcomes were used to develop separate sex-stratified algorithms to sequentially rule out anatomical and functional CAD. We positively labelled patients with a history of ischaemic heart disease, out-of-hospital cardiac arrest, aneurysms, and/or cardiac surgery to reduce the false negative risk.

Modelling and statistical analysis

Continuous variables are presented as means and standard deviations or medians and interquartile ranges and were compared using the unpaired Mann–Whitney U test (non-normal distributed) or Student's t-test (normal distributed). Categorical variables are displayed as frequencies and percentages and compared with χ^2 tests. A two-sided P-value of <0.05 was considered significant.

Outcomes were predicted with XGBoost, an optimized Gradient Boosting algorithm using built-in cross-validation, LASSO (L1), and Ridge (L2) regularization to prevent overfitting.[25] XGBoost is capable of handling sparse data and missing values, which made it suitable for training a model on our EHR data. The original dataset was randomly split into a 90% training set, on which 10-fold cross-validation was performed and a 10% test set. Predictive features were distinguished from the noise features with the Boruta algorithm.[26] Hyperparameter optimization was omitted to avoid the unnecessary risk of overfitting. We did manually set the number of iterations of the gradient boosting based on the validation accuracy. Out-of-fold predictions from 10-fold cross-validation were used for optimization. Because we expected predictive features to differ between men and women, we developed sex-stratified models. For model interpretation, we calculated shapley additive explanations (SHAP) values, where each SHAP value represents the impact that a feature generates in the prediction.

We labelled a patient as true negative if the probability of no CAD was ≥0.95 and the patient had no CAD. The area under the curve (AUC), negative predictive value (NPV), the specificity (SPEC), and false negative rate were assessed as major performance metrics, as we specifically focused on the *absence* CAD on CCTA and SPECT/CMR.

All analyses were done using R 3.6.2, RStudio, and Python 3.9.0.

Results

Population characteristics

The study population comprised of 6808 patients in the CCTA set [33% female; mean age females 54 years (standard deviation, SD 13 years); mean age males 55 years (SD 12 years)], and 3053 patients in the CMR/SPECT set [41% female; mean age females 63 years (SD 12 years); mean age males 62 years (SD 12 years)]. As for the outcome variables, 1092 females (48%) and 1908 males (42%) had no diagnosis of CAD on CCTA, 650 females (47%) and 881 males (45%) had no diagnosis of CAD on SPECT/CMR. Of all patients in the CCTA set, 27% had a history of cardiovascular disease (CVD). Of patients with SPECT/CMR data, 66% had a history of CVD (Table 1).

Table 1   Baseline characteristics CCTA and SPECT/CMR cohorts

	CCTA	SPECT/CMR
N	6808	3381
Age [mean (SD)]	56.03 (12.34)	61.93 (12.14)
Female sex (%)	2253 (33.1)	1394 (41.2)
Hypercholesterolaemia (%)	2719 (39.9)	2324 (68.7)
Hypertension (%)	1084 (15.9)	2016 (59.6)
Diabetes mellitus (%)	654 (9.6)	733 (21.7)
Atrial fibrillation (%)	1280 (18.8)	358 (10.6)
Platelet aggregation inhibitor use (%)		
0	3856 (64.2)	1101 (38.1)
1 or more	2151 (38.5)	1790 (61.9)
Cholesterol (mmol/L) [median (IQR)]	5.20 (4.40–6.10)	4.90 (4.10–5.90)
C-reactive Protein (mg/L) [median (IQR)]	2.70 (1.20–9.30)	4.90 (1.83–33.00)
HDL cholesterol (mmol/L) [median (IQR)]	1.28 (1.07–1.53)	1.26 (1.04–1.53)
LDL cholesterol (mmol/L) [median (IQR)]	3.10 (2.40–3.90)	2.70 (2.10–3.60)
Creatinine (μmol/L) [median (IQR)]	94.00 (80.00–111.00)	92.00 (81.00–106.00)
BMI [median (IQR)]	25.86 (23.41–28.91)	26.40 (23.71–29.45)
Systolic blood pressure (mmHg) [median (IQR)]	135.00 (122.00–150.00)	138.00 (121.00–156.00)
Mean arterial pressure (mmHg) [median (IQR)]	99.00 (91.00–108.00)	97.00 (87.00–107.00)
Diastolic blood pressure (mmHg) [median (IQR)]	81.00 (74.00–89.00)	80.00 (70.00–86.00)

A total of 1421 variables were extracted from UPOD on which the algorithms were trained (Supplementary material online, *Table S1*).

Predicting ≥0.95 probability of having of no coronary artery disease on coronary computed tomography angiography

The CCTA models achieved AUCs of 0.80 (females) and 0.79 (males). Receiver operating characteristic curves are given in *Figure 1*. Of all patients who had no diagnosis of CAD on CCTA, 19% (*n* = 158) (females) and 9% (*n* = 195) (males) could be identified with NPVs of 0.95 (females) and 0.96 (males). False negative rates (type II error) were 0.009 (females) and 0.003 (males), thus, very few patients had predicted probabilities of ≥0.95 of no CAD on CCTA, while CAD on CCTA was observed.

Shapley values reveal that age had the largest impact on the model outcome for both females and males, whereby a lower age reduced the predicted CAD risk (*Figure 2*). For females, missing sodium corresponded to lower CAD risk, as well as not having prescribed platelet aggregation inhibitors, and having a higher glomerular filtration rate. Also, lower mean neutrophil size was associated with lower CAD risk, as well as lower C-reactive protein and lower immature reticulocyte fraction. For males, not having prescribed medication (platelet aggregation inhibitors and medication for the cardiovascular system) was associated with lower CAD risk, as well as not having atrial fibrillation and hypercholesterolaemia. Lower glucose, lower C-reactive protein, and lower neutrophil count corresponded to lower CAD risk. Other features and their impact on the model output can be found in *Figure 2*.

Predicting ≥0.90 probability of having of no coronary artery disease on single-photon emission computerized tomography/cardiac magnetic resonance

The SPECT/CMR models achieved AUCs of 0.61 (females) and 0.60 (males). Receiver operating characteristic curves are given in *Figure 1*. Of all patients who had no diagnosis of CAD on SPECT/CMR, 4% (*n* = 158) (females) and 3% (*n* = 195) (males) could be identified with NPVs of 0.75 (females) and 0.92 (males). False negative rates (type II error) were 0.012 (F) and 0.002 (M), thus, very few patients had predicted probabilities of ≥0.90 of no CAD on SPECT/CMR, while CAD on SPECT/CMR was observed.

For the SPECT/CMR models, less variables had impact on the model output compared to the CCTA models (*Figure 3*). For females, a lower minimum of the ST-segment measured on Lead V4 (5th intercostal space mid clavicular) was associated with lower CAD risk on SPECT/CMR, as well as the non-use of vasodilators and organic nitrates. Lastly, missing HDL cholesterol was predictive of lower CAD risk, as well as lower haemoglobin. For males, the non-use of platelet aggregation inhibitors, a lower S-peak measured on Lead V4 (5th intercostal space mid clavicular), a higher P-wave measured on Lead II, a lower P full area, and a higher ST level at the J point corresponded to lower CAD risk (*Figure 3*).

Discussion

In this EHR cohort study, we trained machine learning models to sequentially exclude anatomical CAD on CCTA and functional CAD on CMR/SPECT in patients with chest discomfort [NPVs for CCTA:
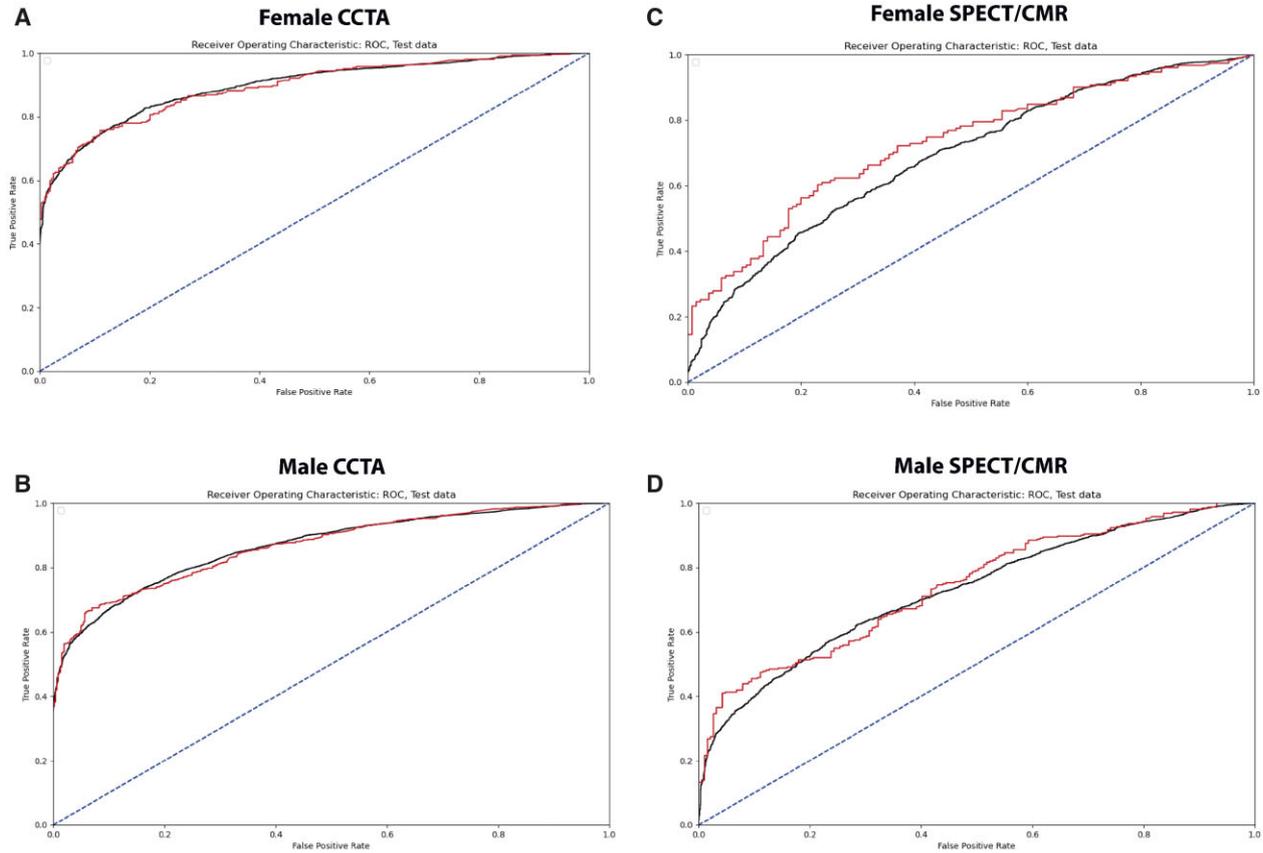
Figure 1 Receiver operating characteristic curves train (black) and test (red) data of the sex-stratified coronary computed tomography angiography and single-photon emission computerized tomography/cardiac magnetic resonance models. (*A*) Receiver operating characteristic curve of the female coronary computed tomography angiography model. (*B*) Receiver operating characteristic curve of the male coronary computed tomography angiography model. (*C*) Receiver operating characteristic curve of the female single-photon emission computerized tomography/cardiac magnetic resonance model. (*D*) Receiver operating characteristic curve of the male single-photon emission computerized tomography/cardiac magnetic resonance model.

0.95 (F) and 0.96 (M), NPVs for CMR/SPECT ≤ 0.73]. Relevant features in the models reflect current care. Furthermore, we demonstrate additional value for haematological biomarkers and raw ECG data to exclude CAD. Lastly, our results confirm the importance of sex-stratification of algorithms in cardiology to capture differences in predictor variables between sexes.

Performance of the SPECT/CMR models is inferior to that of the CCTA models and a probability of no CAD on SPECT ≥0.95 consequently was reached in only few patients. Hence, the threshold for the probability of no CAD on SPECT/CMR had to be lowered to 0.90 for further analyses. Many risk factors for CAD are well established and implemented in clinical care, such as smoking, increased body mass index, and hyperlipidaemia.[27] Therefore, these parameters are routinely assessed by healthcare professionals recorded in the EHR and constitute important features in our models. Evidence on underlying mechanisms of non-obstructive CAD in the absence CAD on CCTA is scarce and not yet included in the clinical cardiology workflow, resulting in less evidence in EHR data to exclude CAD on SPECT/CMR. We found some evidence for the importance of haematological and ECG data, yet these could not fully grasp the disease.

Larger mean neutrophil size and larger immature reticulocyte fraction were important for the exclusion of CAD on CCTA in women. Given the novelty of the haematology data, it can be challenging to interpret these and other haematological predictors biologically. However, previous studies established comparable associations between similar haematological parameters and major adverse cardiovascular events following coronary angiography or the extent of coronary calcifications.[8–11,28] These unique haematology data and raw ECG data allowed us to analyse data beyond traditional risk factors in the EHR and provide opportunities to find new directions for research into CAD.

Our results underline the importance of looking beyond standard cardiovascular modifiable risk factors (SMuRFs) diabetes, hypertension, smoking, and hypercholesterolaemia when making the decision for or against further cardiac imaging in patients presenting with chest discomfort, since SMuRFs may not explain the entire burden of CAD.[29] In fact, the absence of SMuRFs does not guarantee absence of significant CAD. Studies have shown that individuals presenting with ST-elevation myocardial infarction have a significantly increased risk of mortality in the absence of SMuRFS compared to patients with at least one SMuRF, which was found to be particularly applicable to
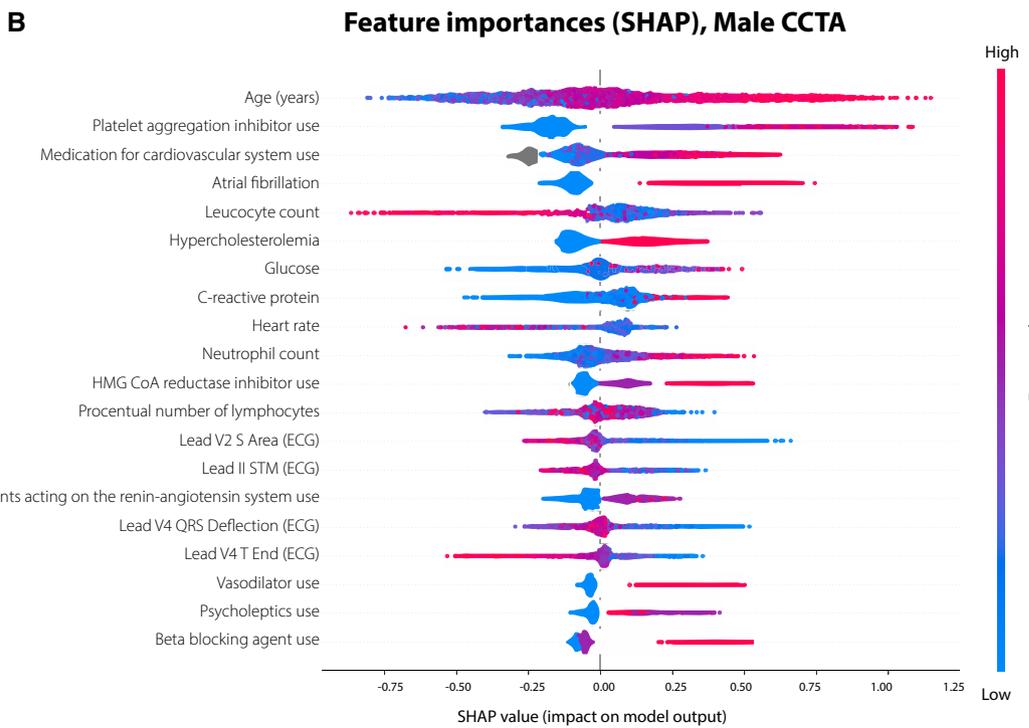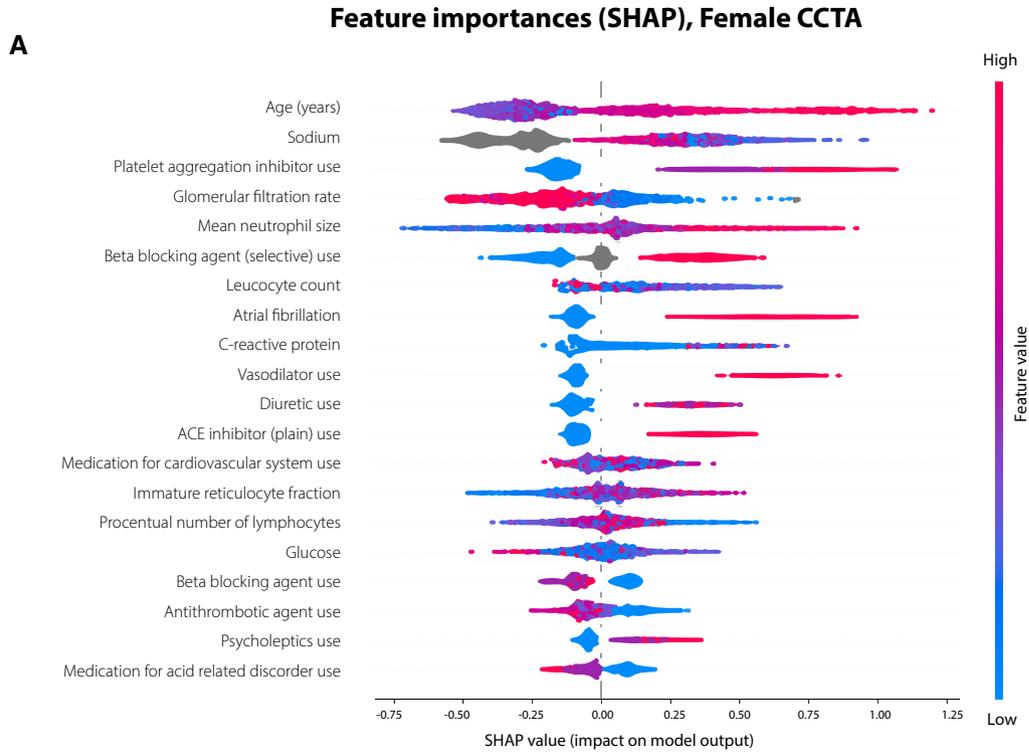
Feature importances (SHAP), Female CCTA

A

Feature importances (SHAP), Male CCTA

B

Figure 2 Shapley summary plots for coronary computed tomography angiography model interpretation. (*A*) Shapley additive explanations values corresponding to the female coronary computed tomography angiography model. (*B*) Shapley additive explanations values corresponding to the male coronary computed tomography angiography model. Features are ranked according to its importance to the model output. The colour represents the value each feature can take, grey for missing values, red for high values, and blue for low values. Low shapley additive explanations values push the model towards class 0 (no coronary artery disease), high shapley additive explanations values push the model output towards class 1 (coronary artery disease).
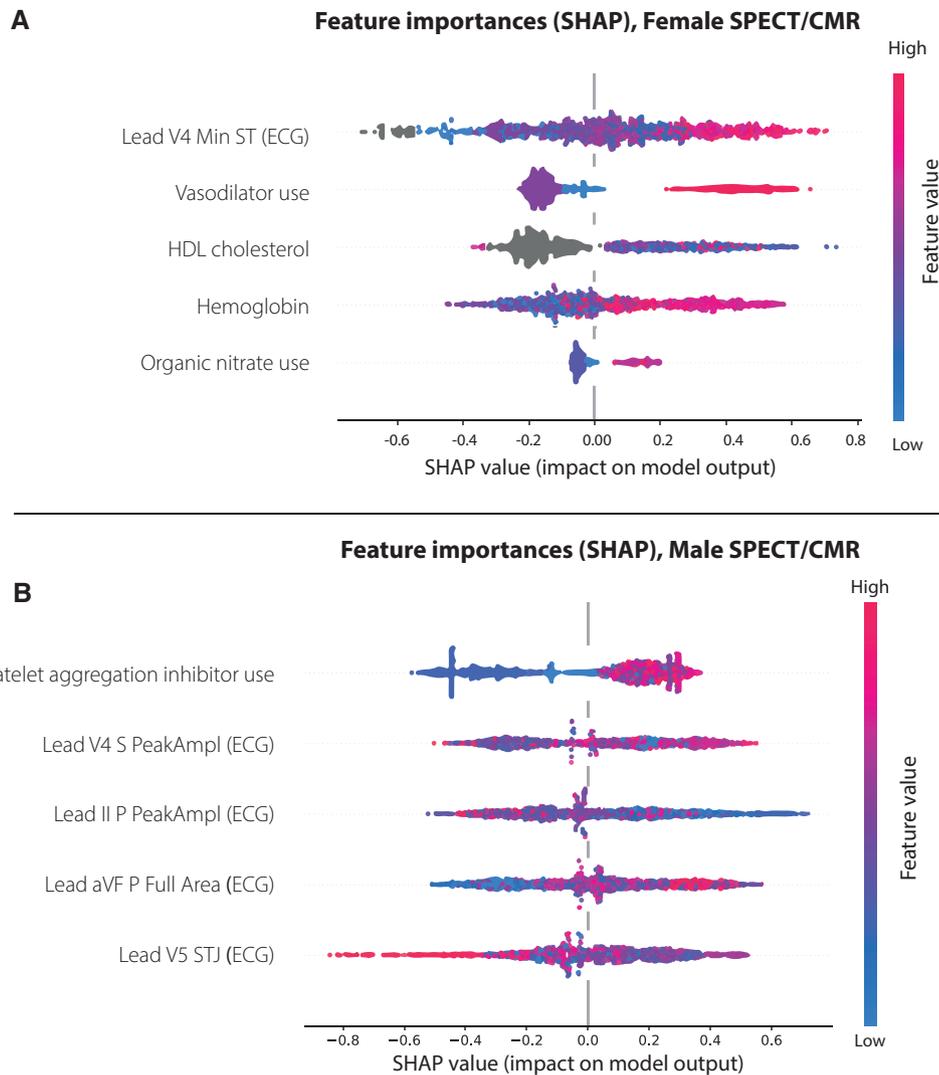
Figure 3 Shapley summary plots for single-photon emission computerized tomography/cardiac magnetic resonance model interpretation. (A) Shapley additive explanations values corresponding to the female single-photon emission computerized tomography/cardiac magnetic resonance model. (B) Shapley additive explanations values corresponding to the male single-photon emission computerized tomography/cardiac magnetic resonance model. Features are ranked according to its importance to the model output. The colour represents the value each feature can take, grey for missing values, red for high values, and blue for low values. Low shapley additive explanations values push the model towards class 0 (no coronary artery disease), high shapley additive explanations values push the model output towards class 1 (coronary artery disease).

women.[30] According to our findings and through new, data-driven techniques, SMuRFs may eventually be expanded to provide even more targeted risk stratification support at the time of referral, as also suggested in other studies.[30,31] However, data-driven approaches such as ours are not suited for determining causal factors and more research is needed on the role of predictor variables in the aetiology of CAD.

A previous study suggested to incorporate the coronary artery calcium score (CACS), next to clinical features, to predict pretest likelihood of CAD on CCTA.[32] However, the CACS this still requires a computed tomography scan to introduce this feature to the machine learning models. Using only features derived from the EHR as they are generated in clinical practice, as in our study, circumvents this.

Other research suggested that the initial Diamond–Forrester model, consisting of age, sex, type of chest pain, overestimated the probability of CAD in a contemporary cohort, especially in women.[33] However, there was no further consideration of functional causes of chest pain in these women, which may cause CAD to be incorrectly excluded. Currently, no model for the exclusion diagnosis of CAD on CCTA and SPECT/CMR exists which uses comprehensive haematological and ECG parameters. Our results suggest that adding these features, if available, can help excluding CAD safely.

This study has several strengths. First, to the best of our knowledge, this is the first large study on EHR data predicting the absence of CAD on both CCTA and SPECT/CMR by routinely available clinical data. This includes both, anatomical coronary abnormalities and

reduced perfusion or myocardial ischaemia, whether caused by significant coronary obstruction or microvascular dysfunction. As our models were made based on readily available data derived from routine care, with no need for additional testing of biomarkers or research parameters, this can help translation into the clinic. Second, we made use of unique and extensive haematological data showing that novel, relatively unexamined parameters are associated with CAD. Third, we decided to include patients with cardiovascular history in our analyses, as in practice, there is no need to distinguish beforehand between subjects with or without prior known CVD. Finally, we applied the recent insights regarding sex differences in the aetiology of cardiovascular disease and our sex-stratified models underline the differences in associated factors between men and women.

Several study limitations should be considered. First, as we trained our models on EHR data of a third-line hospital in the Netherlands, our findings cannot be generalized or directly translated to other centres. It is therefore also important to note that the initial percentage of normal CCTA scans (males 41%, females 48%), and normal SPECT/CMR scans (males 45%, females 48%) was relatively low compared to other reports.[34] This is most likely due to our mentioned third-line hospital population. In many cases, cardiovascular disease had already been excluded at an earlier stage in many low-risk patients before being referred to a third-line centre. Therefore, future work is needed to assess whether our models can be applied in other, low-risk populations. Second, an important characteristic of EHR data is the large number of missing values because of the heterogeneous hospital population and associated protocols. Missing data in EHR however must not be misinterpreted as irrelevant as any clinical consideration leading to not include certain measures or parameters are usually specific for the individual case. Imputing EHR data therefore introduces significant non-random bias and was hence disregarded in our study. Instead, we chose to use the XGBoost algorithm which treats the missing values as an additional category and feature presence as additional parameters as their lack of recording usually also conveys relevant information. Third, due to the design of the study, we are unaware of the specific symptoms of the included patients. Therefore, it is possible that patients were included that are out of scope for the future CDSS. This may cause noise in the training set, reducing the algorithm's performance. Worst-case, this would lead to overestimating the total number of CCTA/SPECT/CMR scans required, which would be performed anyhow without the application of the algorithm.

Conclusions

This study provides promising results for exclusion of anatomical and functional CAD in patients suspect of CAD using EHR data, including extensive haematological data. Built into a CDSS, these algorithms may help guide clinical practice for individual patients and improve the sustainability of the healthcare system.

Supplementary material

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

Data availability

The data underlying this article cannot be shared publicly due to the privacy of patients from whom data have been extracted.

References

1. Population projections in the EU [Internet]. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_projections_in_the_EU (27 September 2021).
2. Kovacic JC, Moreno P, Hachinski V, Nabel EG, Fuster V. Cellular senescence, vascular disease, and aging. *Circulation* 2011;**123**:1650–1660.
3. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. *J Am Coll Cardiol* 1999;**33**:453–462.
4. Shaw LJ, Hachamovitch R, Berman DS, et al. The economic consequences of available diagnostic and prognostic strategies for the evaluation of stable angina patients: an observational assessment of the value of precatheterization ischemia. *J Am Coll Cardiol* 1999;**33**:661–669.
5. Bertoldi EG, Stella SF, Rohde LE, Polanczyk CA. Long-term cost-effectiveness of diagnostic tests for assessing stable chest pain: modeled analysis of anatomical and functional strategies. *Clin Cardiol* 2016;**39**:249–256.
6. Bertoldi EG, Stella SF, Rohde LEP, Polanczyk CA. Cost-effectiveness of anatomical and functional test strategies for stable chest pain: public health perspective from a middle-income country. *BMJ Open* 2017;**7**:e012652.
7. Awaysheh A, Wilcke J, Elvinger F, Rees L, Fan W, Zimmerman KL. Review of medical decision support and machine-learning methods. *Vet Pathol* 2019;**56**:512–525.
8. den Harder AM, de Jong PA, de Groot MCH, et al. Commonly available hematological biomarkers are associated with the extent of coronary calcifications. *Atherosclerosis* 2018;**275**:166–173.
9. Kofink D, Muller SA, Patel RS, et al.; SMART Study Group. Routinely measured hematological parameters and prediction of recurrent vascular events in patients with clinically manifest vascular disease. *PLoS One* 2018;**13**:e0202682.
10. Gijsberts CM, den Ruijter HM, de Kleijn DPV, et al. Hematological parameters improve prediction of mortality and secondary adverse events in coronary angiography patients: a longitudinal cohort study. *Medicine (Baltimore)* 2015;**94**:e1992.
11. Gijsberts CM, Ellenbroek GH, Ten Berg MJ, et al. Routinely analyzed leukocyte characteristics improve prediction of mortality after coronary angiography. *Eur J Prev Cardiol* 2016;**23**:1211–1220.
12. Camici PG, Crea F. Coronary microvascular dysfunction. *N Engl J Med* 2007;**356**:830–840.
13. Jespersen L, Hvelplund A, Abildstrom SZ, et al. Stable angina pectoris with no obstructive coronary artery disease is associated with increased risks of major adverse cardiovascular events. *Eur Heart J* 2012;**33**:734–744.
14. Sharaf B, Wood T, Shaw L, et al. Adverse outcomes among women presenting with signs and symptoms of ischemia and no obstructive coronary artery disease: Findings from the National Heart, Lung, and Blood Institute-sponsored Women's Ischemia Syndrome Evaluation (WISE) angiographic core laboratory. *Am Heart J* 2013;**166**:134–141.
15. von Mering GO, Arant CB, Wessel TR, et al. Abnormal coronary vasomotion as a prognostic indicator of cardiovascular events in women. *Circulation* 2004;**109**:722–725.
16. Reis SE, Holubkov R, Smith AJC, et al. Coronary microvascular dysfunction is highly prevalent in women with chest pain in the absence of coronary artery disease: results from the NHLBI WISE study. *Am Heart J* 2001;**141**:735–741.
17. Christoph M, Christoph A, Dannemann S, et al. Mental symptoms in patients with cardiac symptoms and normal coronary arteries. *Open Heart* 2014;**1**:e000093.
18. Cekirdekci EI, Bugan B. Level of anxiety and depression in cardiac syndrome X. *Med Princ Pract* 2019;**28**:82–86.
19. Piegza M, Pudlo R, Badura-Brzoza K, Hese R. Kardiologiczny zespół X w ujeciu psychosomatycznym [Cardiac syndrome X from a psychosomatic point of view]. *Psychiatr Pol* 2008;**42**:229–236.

20. Conroy RM, Pyorala K, Fitzgerald A. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;**24**: 987–1003.

21. Foldyna B, Udelson JE, Kárády J, et al. Pretest probability for patients with suspected obstructive coronary artery disease: re-evaluating Diamond–Forrester for the contemporary era and clinical implications: insights from the PROMISE trial. *Eur Heart J Cardiovasc Imaging* 2019;**20**:574–581.

22. Groepenhoff F, Eikendal ALM, Bots SH, et al. Cardiovascular imaging of women and men visiting the outpatient clinic with chest pain or discomfort: design and rationale of the ARGUS Study. *BMJ Open* 2020;**10**:e040712.

23. ten Berg MJ, Huisman A, van den Bemt PMLA, Schobben AFAM, Egberts ACG, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med* 2007;**45**:13–19.

24. Groenhof TKJ, Koers LR, Blasse E, et al. Data mining information from electronic health records produced high yield and accuracy for current smoking status. *J Clin Epidemiol* 2020;**118**:100–106.

25. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016*. ACM, San Francisco California USA.

26. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010;**36**:1–13.

27. Khot UN, Khot MB, Bajzer CT. Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA* 2003;**290**:898–904.

28. van Hout GPJ, van Solinge WW, Gijsberts CM, et al. Elevated mean neutrophil volume represents altered neutrophil composition and reflects damage after myocardial infarction. *Basic Res Cardiol* 2015;**110**:58.

29. Vernon ST, Coffey S, Bhindi R, et al. Increasing proportion of ST elevation myocardial infarction patients with coronary atherosclerosis poorly explained by standard modifiable risk factors. *Eur J Prev Cardiol* 2017;**24**:1824–1830.

30. Figtree GA, Vernon ST, Hadziosmanovic N, et al. Mortality in STEMI patients without standard modifiable risk factors: a sex-disaggregated analysis of SWEDEHEART registry data. *The Lancet* 2021;**397**:1085–1094.

31. Figtree GA, Vernon ST, Nicholls SJ. Taking the next steps to implement polygenic risk scoring for improved risk stratification and primary prevention of coronary artery disease. *Eur J Prev Cardiol* 2020;doi:10.1093/eurjpc/zwaa030.

32. Al'Aref SJ, Maliakal G, Singh G, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *Eur Heart J* 2020;**41**:359–367.

33. Genders TSS, Steyerberg EW, Alkadhi H, et al. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. *Eur Heart J* 2011;**32**:1316–1330.

34. Hoffmann U, Ferencik M, Udelson JE, et al. Prognostic value of noninvasive cardiovascular testing in patients with stable chest pain: insights from the PROMISE Trial (Prospective Multicenter Imaging Study for Evaluation of Chest Pain). *Circulation* 2017;**135**:2320–2332.