Title
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.

Permalink
https://escholarship.org/uc/item/7pb3g0r1

Journal
European Heart Journal, 41(3)

ISSN
0195-668X

Authors
Al'Aref, Subhi J
Maliakal, Gabriel
Singh, Gurpreet
et al.

Publication Date
2020-01-14

DOI
10.1093/eurheartj/ehz565

Peer reviewed
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

Aims

Symptom-based pretest probability scores that estimate the likelihood of obstructive coronary artery disease (CAD) in stable chest pain have moderate accuracy. We sought to develop a machine learning (ML) model,
Coronary computed tomography angiography (CCTA) has emerged as an accurate method for the non-invasive evaluation of coronary artery disease (CAD). Numerous studies have shown that the absence of CAD on CCTA conveys very low risk for incident cardiovascular events while there is a graded relationship between the presence of CAD on CCTA and the risk for incident cardiovascular events. Practice guidelines for the management of stable chest pain from the European Society of Cardiology (ESC), American Heart Association (AHA), and American College of Cardiology (ACC) are congruent in their recommendations for the use of CCTA as a primary or secondary diagnostic option in symptomatic individuals deemed to be at an intermediate pretest likelihood of having obstructive CAD. However, in day-to-day clinical practice, a significant number of individuals undergoing CCTA have minimal or no CAD. As a direct consequence of the expanding use of CCTA, there is a growing interest within the medical community regarding ways to optimize patient selection with the goal of improving diagnostic yield and cost-effectiveness of CCTA utilization within the context of clinical practice.

A recent approach has sought to improve risk stratification measures in order to streamline patient selection for CCTA performance. Numerous population-derived risk scores have been developed in order to estimate the pretest likelihood of having CAD, such as the updated Diamond-Forrester (UDF) score. The study population comprised of 13,054 patients, of whom 2380 (18.2%) had obstructive CAD (>50% stenosis). Machine learning with CACS produced the best performance [area under the curve (AUC) of 0.881] compared with ML alone (AUC of 0.773), CAD consortium clinical score (AUC of 0.734), and with CACS (AUC of 0.866) and UDF (AUC of 0.682), P < 0.05 for all comparisons. CACS, age, and gender were the highest ranking features.

Conclusion
A ML model incorporating clinical features in addition to CACS can accurately estimate the pretest likelihood of obstructive CAD on CCTA. In clinical practice, the utilization of such an approach could improve risk stratification and help guide downstream management.

Keywords
Coronary artery disease • Coronary artery calcium score • Machine learning • Coronary computed tomography angiography
history of premature CAD, and baseline cholesterol values were documented.

Coronary computed tomography angiography and coronary artery calcium scanning
Coronary computed tomography angiography image acquisition and processing, as well as coronary artery calcium scanning, were performed in accordance with the guidelines outlined by the Society of Cardiovascular Computed Tomography.17–19 While there were no restrictions in scanner type (single-source, dual-source) or brand (Lightspeed VCT, GE Healthcare, Milwaukee, WI, USA; Somatom Definition CT, Siemens, Erlangen, Germany), all machines were required to pass a minimum of 64-detector rows. Level III-equivalent readers evaluated all patient scans and determined the extent of CAD in addition to providing a CACS using the Agatston method. Such a method is semi-automated to calculate a weighted sum of the area of coronary calcification, wherein each calcified area is multiplied by a local density factor determined by the Hounsfield unit (HU) of the calcium (0: 0–129 HU; 1: 130–199 HU; 2: 200–299 HU; 3: 300–399 HU; 4: >400 HU). The outcome of the present study was the presence of obstructive CAD on CCTA, defined as the detection of ≥50% diameter stenosis in any of the four major epicardial coronary arteries. A sensitivity analysis was further performed with the definition of obstructive CAD set at ≥70% diameter stenosis.

Machine learning
Patients included in our analyses were characterized by a total of 25 readily available demographic and clinical variables, including age, gender, risk factors (including diabetes mellitus, hypertension, dyslipidaemia), and baseline cholesterol levels (including total cholesterol, LDL and HDL values). Correlation coefficients between variables were obtained and are shown in Supplementary material online, Figure S1. An ensemble ML algorithm was constructed to classify patients on the basis of the presence of obstructive CAD. Machine learning techniques were implemented in Python using open-source libraries. A gradient boosting machine learning algorithm (XGBoost) was employed for a binary classification task based on the presence or absence of obstructive CAD. XGBoost is a novel boosting tree-based ensemble algorithm which has gained wide popularity in the ML community. XGBoost outlines the creation of classification and regression trees, in which classification accuracy is iteratively improved one level at a time through optimization of a customized objective function—an instance of a process otherwise known as ‘boosting’.20 This algorithm was employed due to its state-of-the-art accuracy; ability to employ both continuous and categorical inputs, without need for scaling or other pre-processing modifications; capacity for handling of sparsity; interpretability; and lastly, high degree of internal optimization and relatively modest computational cost. Overall, the original dataset was randomly split into training (75%) and a held-out validation (25%) set, such that the ratio of obstructive to non-obstructive CAD was maintained across both the training and validation subsets. Further, the training set was divided into 10 equally sized folds roughly maintaining the ratio of event to non-events seen in the training set to select optimal model hyper-parameters by grid search through 10-fold cross-validation. Model hyper-parameters (e.g. number of trees, depth of each tree) were fine-tuned using 10-fold cross-validation on the training set. Cross-validation is an iterative process whereby the training data is partitioned into roughly equally sized subsets (e.g. 10 such subsets in 10-fold cross-validation), with training occurring using all but one of these subsets and validation being performed on that which is remaining. Employing such a tactic during the training phase can be beneficial for numerous reasons which have been well-characterized elsewhere, but its primary use is in empirically determining optimal model hyper-parameters without recourse to the validation set.21 Finally, classification performance of the ML model was measured using the area under the curve (AUC) and the associated 95% confidence interval (CI) and reported for the held-out validation set. Feature ranking was obtained by computing Shapley Additive Explanation values (SHAP), as previously described.22

Statistical analysis
Performance of the ML model to classify participants was compared with commonly employed prediction scores such as the CAD consortium clinical score and the updated Diamond-Forrester (UDF) score. Further, CACS was added to the ML model and the CAD consortium clinical score given its widespread use as a screening modality. Calibration of the ML model (with and without CACS) was evaluated using the calibration slope and the Brier score (on a scale ranging from 0 to 1). The calibration slope was obtained by fitting a linear regression equation on the mean predicted probabilities vs. fraction of positives, and calculating its slope. The Brier score, on the other hand, calculates the difference between the estimated and observed risk for occurrence of obstructive CAD, with values closer to 0 indicating better calibration. Continuous net reclassification index (NRI) was performed in order to quantify how well the ML models reclassified subjects, either appropriately or inappropriately, compared with the traditional UDF score and the CAD consortium clinical score. Finally, continuous variables were expressed as the mean ± 1 SD, while categorical variables were expressed as counts (percentages) of the total population. Comparisons were considered statistically significant based on a two-sided P-value of <0.05.

Results
A total of 13,054 patients met the inclusion criteria and were included in the analysis. The occurrence of obstructive CAD was 18.2% (2380/13,054) within the studied cohort. Mean age 58.0 ± 11.4 and 54% were male patients. Hypertension (52.6%) and hyperlipidaemia (56.9%) were the two prevalent risk factors, while diabetes mellitus was present in 14.2% of the population and 17.4% were

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>58.0 ± 11.4</td>
</tr>
<tr>
<td>Male participants (%)</td>
<td>54%</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>52.9%</td>
</tr>
<tr>
<td>Hyperlipidaemia (%)</td>
<td>56.9%</td>
</tr>
<tr>
<td>Diabetes mellitus (%)</td>
<td>14.2%</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>17.4%</td>
</tr>
<tr>
<td>Mean serum creatinine (mg/dL)</td>
<td>0.97 ± 0.7</td>
</tr>
<tr>
<td>Mean total cholesterol (mg/dL)</td>
<td>189.1 ± 43.6</td>
</tr>
<tr>
<td>Low density lipoprotein (mg/dL)</td>
<td>115.7 ± 36.8</td>
</tr>
<tr>
<td>High density lipoprotein (mg/dL)</td>
<td>52.2 ± 16.0</td>
</tr>
<tr>
<td>Chest pain (%)</td>
<td>69.5%</td>
</tr>
<tr>
<td>Shortness of breath (%)</td>
<td>21.0%</td>
</tr>
</tbody>
</table>
active smokers. At the time of enrolment, mean total cholesterol (in mg/dL) was 189.1 ± 43.6, mean LDL was 115.7 ± 36.8 and mean HDL was 52.2 ± 16.0. Of the total cohort, 68.0% had chest pain while 21.0% had shortness of breath as the presenting symptoms (Table 1).

Prediction of obstructive coronary artery disease on coronary computed tomography angiography

Machine learning produced the best performance in terms of predicting individuals with obstructive CAD, with an AUC of 0.773 (95% CI 0.757–0.791) compared with CAD consortium clinical score (AUC 0.734, 95% CI 0.717–0.751), and UDF score (AUC 0.682, 95% CI 0.662–0.702), P < 0.05 for all comparisons. With the addition of CACS, both ML model and CAD consortium clinical scores significantly improved in terms of prediction of the occurrence of obstructive vs. non-obstructive CAD (AUC 0.881, 95% CI 0.869–0.895 and an AUC 0.866, 95% CI 0.852–0.879, respectively) (Figure 1). Sensitivity analysis was performed with obstructive CAD defined as ≥70% diameter stenosis with no significant change in discriminative performance across the five models (Supplementary material online, Figure S2). The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for the prediction of obstructive CAD (at a probability threshold of 15%) were 78.0%, 62.8%, 31.9%, 92.8%, and 81.3% for Model 1 (ML) and 80.0%, 81.5%, 49.1%, 94.8%, and 81.3% for Model 2 (ML + CACS), respectively.

Figure 1  Area under the curve as a measure of individual model performance for the prediction of obstructive coronary artery disease on coronary computed tomography angiography. AUC, area under the curve; CACS, coronary artery calcium score; CAD, coronary artery disease; ML, machine learning; UDF, updated Diamond-Forrester score.

Figure 2  Feature importance plot for the (A) machine learning model and (B) machine learning model with coronary artery calcium score. The top 20 clinical variables are shown in this figure. The blue and red points in each row represent participants having low to high values of the specific variable, while the x-axis gives the SHAP value which gives the impact on the model (i.e. does it tend to drive the predictions towards event (positive value of SHAP) or non-event (negative value of SHAP)). SHAP, Shapley Additive Explanation values.
The performance of the ML model was subsequently evaluated in select subgroups stratified by age, gender, presence of diabetes mellitus and/or chest pain typicality. The ML model showed improved discrimination for the detection of obstructive CAD in younger individuals (less than 65 years of age) with atypical chest pain (Supplementary material online, Figure S3). Interestingly, there appear to be gender-specific differences in the performance of the ML model since the ML model with CACS is better in males with atypical chest pain than in females with atypical chest pain.

**Predictive features**

As shown in Figure 2, age, ethnicity, and sex were the most predictive features in the ML model, followed by prior history of hypertension and hypercholesterolaemia. Interestingly, with the addition of CACS into the model, the most predictive features (after the CACS itself) were age and sex followed by history of cerebrovascular disease and the presence of shortness of breath as the presenting symptom. In both models, the presence of shortness of breath was more predictive of the presence of obstructive CAD than the presence of chest pain and typicality of symptoms. Coronary artery calcium score had the highest predictive value as low CACS values were likely to be associated with absence of obstructive CAD (blue colour), while very high values (red colour) were significantly associated with obstructive CAD (Figure 2B).

**Calibration and net reclassification**

Model calibration was performed in order to assess the certainty of a given new observation belonging to each of the already established classes (prediction of the presence or absence of CAD on CCTA). The calibration slope was 0.856 for the ML model and 0.992 after the addition of CACS, indicating minimal difference between the predicted and observed probability of obstructive CAD and hence good model fit (Figure 3). On the other hand, the Brier score for Model 1 (ML) was 0.205 before and 0.127 after calibration. Similarly to Model 1, Model 2 (ML with CACS) had a Brier score of 0.224 before and a score of 0.099 after calibration. Additionally, continuous NRI was performed comparing both ML models (Models 1 and 2) to the conventional comparator risk scores (Models 3, 4, and 5). Net reclassification index was 0.585 (95% CI 0.495–0.671) when Model 1 (ML)
was compared with Model 3 (CAD consortium) and an NRI of 0.685 (95% CI 0.600–0.770) when compared with Model 5 (UDF). Similarly, Model 2 performed better than Model 4 (CAD consortium + CAC score: NRI 0.816; 95% CI 0.713–0.900) and Model 5 (UDF: NRI of 1.144; 95% CI 1.067–1.220) (Figure 4). All NRIs were driven both by an improvement in correct classification of both events (i.e. correct prediction of obstructive CAD) and non-events (i.e. correct prediction of non-obstructive CAD).

**Discussion**

Coronary artery disease is a commonly encountered disease entity associated with significant morbidity, mortality, and healthcare expenditure. A conventional routine in clinical practice over the years has been to employ validated diagnostic models of the PTP of stable, albeit obstructive, CAD in order to direct downstream testing. A majority of existent models have modest performance (with remarkable overestimation of risk in certain subgroups such as women) while very few studies have data regarding the effect of PTP-based models on clinical decision-making regarding further testing or patient outcomes. Hence, there is a need for clinically based models that can predict the PTP of stable CAD and as a result function as gatekeepers to identify low-risk individuals who are unlikely to have obstructive CAD and unlikely to need further diagnostic testing. In the present investigation, we utilized readily available clinically characteristic in a large multicentre, multiethnic cohort undergoing clinically indicated CCTA for the diagnosis of CAD. We utilized ML as a novel analytic approach that is optimized towards the creation of accurate predictive models and found that the developed ML model predicts the occurrence of obstructive CAD on CCTA, specifically in younger individuals with atypical symptoms. Added to that was the finding of

---

**Take home figure**  The developed machine learning model incorporates readily available clinical characteristics and improves the ability to rule-out the presence of obstructive coronary artery disease. This approach could improve decision-making and streamline resources to the appropriate risk individuals.
cognition of obstructive CAD compared with the DCS by itself (AUC 0.79 to 0.88 for the prediction of obstructive CAD on CCTA compared with 44.5% on single-photon emission computed tomography and 43.8% on stress echocardiography).\cite{23,24} The ML + CACS approach as a routine in symptomatic patients with an intermediate likelihood of obstructive CAD could lead to improved patient outcomes as well as to guide downstream testing. The dramatic reduction in myocardial infarction and cardiac death noted in the SCOT-Heart trial has been attributed to both institution of preventive therapies as well as the use of early revascularization. Current guidelines suggest that CACS can be used to guide the use of preventive therapies in asymptomatic patients at intermediate risk. In clinical practice, it is also used to guide the use of these therapies in symptomatic patients. For patients in whom the ML + CACS were to fall below the level that indicated the need for further testing, preventive therapy resulting from the combined score might lead to improved outcomes. Similarly, if the result of the ML + CACS prompted the use of stress testing, the CACS could lead to appropriate preventive therapies in high proportion of patients in whom the stress testing is negative.

An important consideration for the use of ‘big data’ for predictive modelling within clinical practice is the need for existence of standardization as well as quality control measures for acquisition and processing of diagnostic imaging results. In the example of CCTA, there exist numerous image acquisition and processing protocols between various sites and institutions. Such variability can be attributed to both hardware and software, as well as variations in interpretation and reporting of imaging findings. There have been several studies that evaluated the influence of intra- and inter-scanner variability on cardiovascular findings on CCTA. As such, the creation of standardized systems for both image acquisition and assessment of findings is of paramount importance especially for the creation of useful and reproducible imaging-based prediction scores. Furthermore, the integration of such risk assessment tools into electronic health record systems would permit continuous optimization of scores at a system level and thereafter help guide with clinical decision-making. Incorporation of a multitude of clinical systems, such as qualitative and quantitative imaging findings, ‘omic’ data such as genomics and proteomics, will advance abilities further, as will data capture from novel technology, such as wearable gadgets.
There are several limitations of the present investigation that are noteworthy to mention. Firstly, the CONFIRM cohort includes participants referred for CCTA for the evaluation of suspected CAD. As a result of the limitations associated with such a design, there could be a significant referral bias that results in a predictive model that does not apply to a community or general cohort of individuals. However, the goal of utilizing ML was to introduce the concept that such an approach is specifically tailored towards big data, wherein model parameters can be automatically updated and recalibrated as more data becomes available. Secondly, external validation in an independent cohort was not done in the present investigation but is planned for a subsequent analysis on well-validated cohorts of stable chest pain such as PROMISE and SCOT-HEART. Thirdly, UDF and CAD clinical consortium scores were available as comparator scores, while other commonly utilized stratification scores, such as the DCS, were not available for comparison. Nevertheless, the present study included the largest cohort, to date, used for the development of a predictive score for the presence of obstructive CAD on CCTA from a multicenter and multiethnic cohort. Additionally, all included participants had the outcome of interest (i.e. determination of CAD severity on CCTA) without the need to apply imputation techniques. We acknowledge that in the presence of severe calcification (i.e. a high CACS), CCTA overestimates % stenosis, hence our study endpoint (>50% stenosis by CCTA) does not reflect the effective >50% stenosis by coronary angiography. Given the increasing overestimation of degree of % stenosis by CCTA along with increasing CACS, it is not surprising that Model 2 (ML + CACS) outperformed Model 1 (ML). We recognize that using >50% stenosis by coronary angiography as endpoint, different results may be found.

In conclusion, we developed a ML model based on baseline demographic and clinical characteristics for the prediction of obstructive CAD on CCTA that is highly accurate and results in improved net reclassification as a result of correct reclassification of both obstructive and non-obstructive CAD on CCTA. Additionally, the incorporation of CACS further improves risk stratification, such that an ML score that incorporates the CACS and clinical variables may be optimal for initial assessment of younger individuals with atypical symptoms. The utilization of such models may improve decisions in low to intermediate risk patients regarding the need for further testing such as CCTA, as well as for the need for preventive therapies.

Supplementary material

Supplementary material is available at European Heart Journal online.

Funding

The research reported in this publication was funded, in part, by the National Institute of Health (Bethesda, MD, USA) under award number R01 HL115150. This research was also supported, in part, by the Dalio Institute of Cardiovascular Imaging (New York, NY, USA) and the Michael Wolk Foundation (New York, NY, USA).

Conflict of interest: All authors have completed the ICMJE uniform disclosure form and declare the following: J.K.M. receives funding from the Dalio Foundation, National Institutes of Health, and GE Healthcare, serves on the scientific advisory board of Arineta and GE Healthcare and has an equity interest in Cleerey. Matthew Budoff receives grant support from the National Institutes of Health and General Electric. D.S.B. receives a research grant from Heartflow and software royalties from Cedars-Sinai Medical Center. Gianluca Pontone has a research grant and/or honorarium as speaker from GE, Bracco, Bayer, Medtronic, and Heartflow. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

References


Results from a multicenter, statewide registry, the Advanced Cardiovascular Imaging Consortium. J Am Coll Cardiol 2012;60:1185–1191.


