

Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis

Manish Motwani¹, Damini Dey¹, Daniel S. Berman¹, Guido Germano¹, Stephan Achenbach², Mouaz H. Al-Mallah³, Daniele Andreini⁴, Matthew J. Budoff⁵, Filippo Cademartiri^{6,7}, Tracy Q. Callister⁸, Hyuk-Jae Chang⁹, Kavitha Chinnaiyan¹⁰, Benjamin J.W. Chow¹¹, Ricardo C. Cury¹², Augustin Delago¹³, Millie Gomez¹⁴, Heidi Gransar¹, Martin Hadamitzky¹⁵, Joerg Hausleiter¹⁶, Niree Hindoyan¹⁴, Gudrun Feuchtner¹⁷, Philipp A. Kaufmann¹⁸, Yong-Jin Kim¹⁹, Jonathon Leipsic²⁰, Fay Y. Lin¹⁴, Erica Maffei²¹, Hugo Marques²², Gianluca Pontone²³, Gilbert Raff¹⁰, Ronen Rubinshtein²⁴, Leslee J. Shaw²⁵, Julia Stehli¹⁸, Todd C. Villines²⁶, Allison Dunning²⁷, James K. Min²⁸, and Piotr J. Slomka^{1*}

¹Departments of Imaging and Medicine and the Cedars-Sinai Heart Institute, and the Biomedical Imaging Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ²Department of Cardiology, Friedrich Alexander Universität Erlangen-Nürnberg, Germany; ³King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center, King AbdulAziz Cardiac Center Saudia Arabia; ⁴Department of Clinical Sciences and Community Health, University of Milan, Centro Cardiologico Monzino, IRCCS, Milan, Italy; ⁵Department of Medicine, Harbor UCLA Medical Center, Los Angeles, CA, USA; ⁶Department of Radiology, Montréal Heart Institute/Université de Montréal, Montréal, Quebec, Canada; ⁷Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands; ⁸Tennessee Heart and Vascular Institute, Hendersonville, TN, USA; ⁹Division of Cardiology, Severance Cardiovascular Hospital, Seoul, South Korea; ¹⁰William Beaumont Hospital, Royal Oaks, MI, USA; ¹¹Department of Medicine and Radiology, University of Ottawa Heart Institute, ON, Canada; ¹²Miami Cardiac and Vascular Institute, Miami, FL, USA; ¹³Capitol Cardiology Associates, Albany, NY, USA; ¹⁴Dalio Institute of Cardiovascular Imaging, Weill Cornell Medical College and New York-Presbyterian Hospital, New York, NY, USA; ¹⁵Department of Radiology and Nuclear Medicine, German Heart Center Munich, Munich, Germany; ¹⁶Medizinische Klinik I der Ludwig-Maximilians-Universität München, Munich, Germany; ¹⁷Department of Radiology, Medical University of Innsbruck, Innsbruck, Austria; ¹⁸Department of Nuclear Medicine, Cardiac Imaging, University Hospital Zurich, Zurich, Switzerland; ¹⁹Department of Radiology and Nuclear Medicine, Seoul National University Hospital, Seoul, South Korea; ²⁰Department of Medical Imaging and Division of Cardiology, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada; ²¹Montréal Heart Institute, Montréal, Quebec, Canada; ²²Department of Radiology, UNICA - Hospital da LUZ, Lisbon, Portugal; ²³Centro Cardiologico Monzino, IRCCS, Milan, Italy; ²⁴Department of Cardiology at the Lady Davis Carmel Medical Center, The Ruth and Bruce Rappaport School of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; ²⁵Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA; ²⁶Department of Medicine, Walter Reed National Medical Center, Bethesda, MD, USA; ²⁷Duke Clinical Research Institute, Durham, NC, USA; and ²⁸Departments of Radiology and Medicine, Weill Cornell Medical College, Dalio Institute of Cardiovascular Imaging, New York-Presbyterian Hospital, New York, NY, USA

Online publish-ahead-of-print 1 June 2016

See page 508 for the editorial comment on this article (doi:10.1093/eurheartj/ehw217)

Aims

Traditional prognostic risk assessment in patients undergoing non-invasive imaging is based upon a limited selection of clinical and imaging findings. Machine learning (ML) can consider a greater number and complexity of variables. Therefore, we investigated the feasibility and accuracy of ML to predict 5-year all-cause mortality (ACM) in patients undergoing coronary computed tomographic angiography (CCTA), and compared the performance to existing clinical or CCTA metrics.

Methods and results

The analysis included 10 030 patients with suspected coronary artery disease and 5-year follow-up from the COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter registry. All patients underwent CCTA as their standard of care. Twenty-five clinical and 44 CCTA parameters were evaluated, including segment stenosis score (SSS), segment involvement score (SIS), modified Duke index (DI), number of segments with non-calcified, mixed or calcified plaques, age, sex, gender, standard cardiovascular risk factors, and Framingham risk score (FRS). Machine learning involved automated feature selection by information gain ranking, model building with a boosted ensemble algorithm, and 10-fold stratified cross-validation. Seven hundred and forty-five patients died during 5-year follow-up.

* Corresponding author. Artificial Intelligence in Medicine Program, Departments of Imaging and Medicine, and Cedars-Sinai Heart Institute, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, A047N, Los Angeles, CA 90048, USA. Tel: +1 310 423 4348, Email: piotr.slomka@cshs.org

Machine learning exhibited a higher area-under-curve compared with the FRS or CCTA severity scores alone (SSS, SIS, DI) for predicting all-cause mortality (ML: 0.79 vs. FRS: 0.61, SSS: 0.64, SIS: 0.64, DI: 0.62; $P < 0.001$).

Conclusions

Machine learning combining clinical and CCTA data was found to predict 5-year ACM significantly better than existing clinical or CCTA metrics alone.

Keywords

Coronary artery disease • Coronary CT angiography • Prognosis • Machine learning

Introduction

Coronary computed tomography angiography (CCTA) is an accurate non-invasive technique for the diagnosis and exclusion of obstructive coronary artery disease (CAD).¹ In addition to coronary stenosis, CCTA also allows for evaluation of coronary atherosclerosis extent, severity, distribution, and composition. Beyond clinical variables, these imaging findings add incremental utility for prediction of future adverse events.^{2–4}

Machine learning (ML) is a field of computer science that uses computer algorithms to identify patterns in large datasets with a multitude of variables, and can be used to predict various outcomes based on the data. Machine learning algorithms typically build a model from test inputs in order to make data-driven predictions or decisions. In recent years, ML techniques have emerged as highly effective methods for prediction and decision-making in a multitude of disciplines, including internet search engines, customized advertising, natural language processing, finance trending, and robotics.^{5,6} There is consensus in the ML community on the superior performance of aggregation approaches, known as ensemble methods such as ‘boosting’, and therefore we have chosen a boosting approach for this study.⁷ To date, the benefit of utilizing ML with data from CCTA to predict hard-prognostic endpoints has not been evaluated on a large scale. Therefore, the aim of this study was to utilize a large, prospective multinational registry of patients undergoing CCTA to assess the feasibility and accuracy of ML to predict 5-year all-cause mortality (ACM), and then compare the performance to existing clinical or CCTA metrics.

Methods

Study population

Ten thousand and thirty stable patients with suspected CAD and with 5-year follow-up from the COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter (CONFIRM) registry were studied.^{2,8} COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter is an international, multicentre, observational registry collecting clinical, procedural, and follow-up data of patients undergoing clinically indicated CCTA (enrolled 2004–10).² The study complies with the Declaration of Helsinki, all patients provided informed consent for inclusion in the registry, and institutional review board approval was obtained at each centre. Individuals with known CAD (defined as prior myocardial infarction (MI) or revascularization) or those with early revascularization after the index CCTA (defined as within 90 days) were excluded from the present study.

Clinical data

A structured interview was conducted before CCTA to collect information on symptoms and the presence of cardiovascular risk factors

(Figure 1). Hypertension was defined as a history of blood pressure > 140 mmHg or treatment with antihypertensive medications. Diabetes mellitus was defined by a diagnosis made previously by a physician and/or use of insulin or oral hypoglycaemic agents. Smoking history was defined as current smoking, or cessation of smoking within the last 3 months. Family history of premature CAD was defined as MI in a first-degree relative < 55 years (male) or < 65 years (female). Dyslipidaemia was defined as known but untreated dyslipidaemia, or current treatment with lipid-lowering medications. In addition, blood cholesterol levels of the lipid test nearest to the index examination were recorded. The primary outcome of the study was ACM recorded as a binary variable, and the study end-point was time to ACM. In U.S. sites, death status was ascertained by querying the Social Security Death Index. In non-U.S. sites, follow-up data were collected by mail or telephone contact with the patients or their families; events were verified by hospital records or contacts with the attending physician.

Coronary computed tomographic angiography data

Image acquisition

Coronary computed tomographic angiography investigations were performed on ≥ 64 -detector row scanners from a variety of vendors (Lightspeed VCT, GE Healthcare, Milwaukee, WI, USA; Somatom Definition CT, Siemens, Erlangen, Germany). The imaging protocol adhered to Society of Cardiovascular Computed Tomography guidelines.⁹ Detailed methodology has been previously published.²

Image analysis

All scans were analysed by level III-equivalent cardiologists or radiologists. Images were evaluated by an array of post-processing techniques, including axial, multi-planar reformat, maximum intensity projection, and short-axis cross-sectional views.

Coronary segments were scored visually for the presence and composition of coronary plaque, and degree of luminal stenosis using a 16-segment coronary artery model.⁹ In each segment, plaques were classified as non-calcified, mixed, or calcified. The presence of coronary calcification was determined visually in contrast-enhanced datasets. Non-calcified plaque was defined as a tissue structure > 1 mm² that could be clearly discriminated from the vessel lumen and surrounding tissue, with a density below the blood-pool. Plaques meeting this definition and in addition showing any calcified areas were classified as mixed. Severity of luminal stenosis in each segment was scored visually: 0 (none, 0%), 1 (mild, 1–49%), 2 (moderate, 50–69%), or 3 (severe, $\geq 70\%$).

For purposes of classification for per-vessel analyses, we considered four arterial territories: left main (LM), left anterior descending (LAD) and its diagonal branches, left circumflex (LCx) and its obtuse marginal branches, and the right coronary artery (RCA). The posterior descending artery and posterolateral branch were considered as part of the RCA or LCx system, depending on recorded dominance.

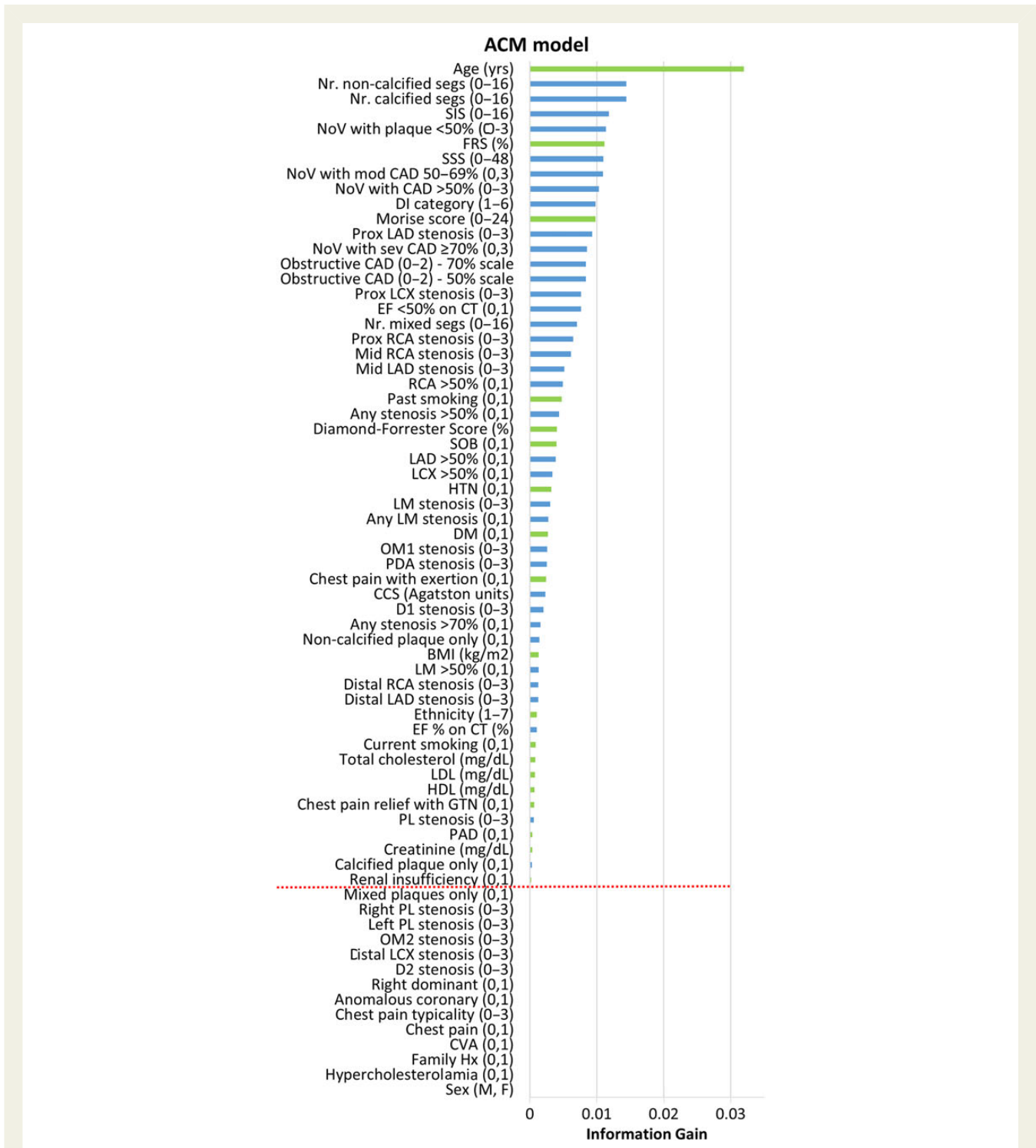


Figure 1 Feature selection. Forty-four coronary computed tomographic angiography variables (blue) and 25 clinical variables (green) were available. Information gain ranking was used to evaluate the worth of each variable by measuring the entropy gain with respect to the outcome, and then rank the attributes by their individual evaluations (top to bottom). Only attributes resulting in information gain >0 (above red line) were subsequently used in boosting. This figure shows the results from one representative fold of the cross-validation procedure. Variables are followed by units, or categorical range in parentheses—full details in Supplementary material online, Appendix. ACM, all-cause mortality; BMI, body mass index; CAD, coronary artery disease; CCS, coronary calcium score; CVA, cerebrovascular accident; D, diagonal; DM, diabetes mellitus; EF, ejection fraction; F, female; FRS, Framingham risk score; HDL, high-density lipoprotein; HTN, hypertension; LM, left main; LAD, left anterior descending artery; LCX, left circumflex; LDL, low-density lipoprotein M, male; NoV, number of vessels; Nr., number of; OM, obtuse marginal; PAD, peripheral arterial disease; PL, posterolateral branch; RCA, right coronary artery; sev, severe; SSS, segment stenosis score; SIS, segment involvement score.

Comparative metrics

Clinical and CCTA parameters were summarized with a number of existing metrics to allow direct comparison with the ML method.

Clinical data

Framingham risk score (FRS) was used as a comparative summary metric for clinical parameters. Framingham risk score was calculated with the established categorical model using low-density lipoprotein cholesterol according to Wilson *et al.*¹⁰ Although, FRS was originally validated for the 10-year risk of coronary heart disease in asymptomatic patients, this is a commonly adopted strategy given the absence of any more suitable clinical risk score for symptomatic patients or shorter follow-up periods. Moreover, it frames the accuracy of the ML-model in the context of a widely used and understood clinical score.^{2–4}

Coronary computed tomographic angiography data

Three composite CCTA-based scores assessing overall plaque burden and severity of CAD were derived. First, the segment stenosis score (SSS) was employed as an overall measure of coronary plaque extent.² For each patient, individual coronary segments were scored 0–3 based on luminal diameter stenosis, and then summed to yield a total SSS (0–48). Second, the segment involvement score (SIS) was calculated as a measure of overall coronary plaque distribution, by summation of the absolute number of coronary segments with plaque (0–16).² Third, the modified Duke prognostic CAD index (DI) was used, which categorizes patients into the following subsets, each with an increasing risk of 5-year death: (1) <50% stenosis, (2) ≥ 2 stenoses 30–49% (including 1 artery with proximal disease) or 1 vessel with 50–69% stenosis, (3) 2 stenoses 50–69% or 1 vessel with $\geq 70%$ stenosis, (4) 3 stenoses 50–69% or 2 vessels with $\geq 70%$ stenosis or proximal left anterior descending stenosis $\geq 70%$, (5) 3 vessels $\geq 70%$ stenoses or 2 vessels $\geq 70%$ stenosis with proximal LAD, and (6) LM stenosis $\geq 50%$.²

Machine learning

Forty-four CCTA parameters and 25 clinical parameters were available (Figure 1). Machine learning involved automated feature selection by information gain ranking, model building with a boosted ensemble algorithm, and 10-fold stratified cross-validation for the entire process (Figure 2).^{11–14} Machine learning techniques were implemented in the open-source Waikato Environment for Knowledge Analysis platform (3.7.12).^{14,15}

Feature selection

Feature selection was performed using a technique known as ‘information gain attribute ranking’.¹¹ Information gain is defined as a measure of the effectiveness of an attribute in classifying the training data. It is measured as the amount by which the entropy of the class decreases, which reflects the additional information about the class provided by the attribute. Only attributes resulting in information gain > 0 were subsequently used in model building (Figure 1).

Model building

Predictive classifiers for ACM prediction were developed by an ensemble classification approach (‘boosting’), employing an iterative LogitBoost algorithm using decision stumps (single-node decision trees) for each feature-selected variable as base classifiers.^{12,16} The principle behind ML ensemble boosting is that a set of weak base classifiers can be combined to create a single strong classifier by iteratively adjusting their appropriate weighting according to misclassifications. A series of base classifier predictions and an updated weighting distribution are produced with each iteration. These predictions are then combined

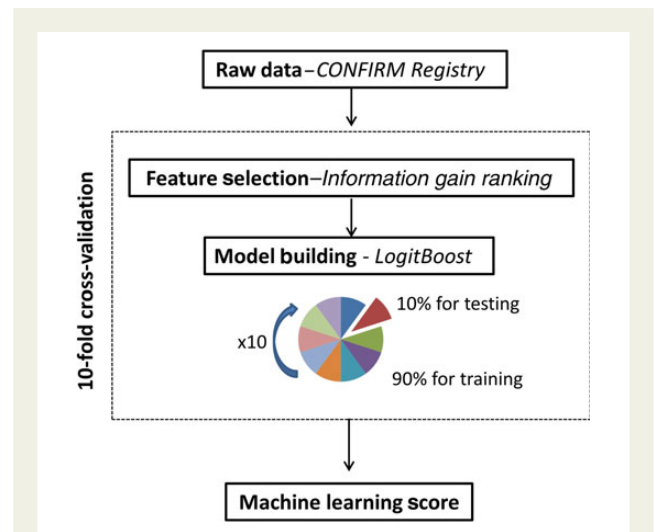


Figure 2 Computational methods. Machine learning involved automated feature selection by information gain ranking, model building with a boosted ensemble algorithm (LogitBoost), and 10-fold stratified cross-validation.

by weighted majority voting to derive an overall classifier—the ‘LogitBoost score’ or risk probability estimate (%).

Cross-validation

The performance and general error estimation of the entire ML process (feature selection and LogitBoost) was assessed using stratified 10-fold cross-validation which is currently the preferred technique in data mining.^{13,17} The main advantages of this validation technique compared with the conventional split-sample (test and validation) approach are: (i) it reduces the variance in prediction error leading to a more accurate estimate of model prediction performance; (ii) it maximizes the use of data for both training and validation, without overfitting or overlap between test and validation data; and (iii) guards against testing hypotheses suggested by arbitrarily split data (Type III errors).¹³

Briefly, the dataset is randomly divided into 10 equal folds, each with approximately the same number of events; 10 validation experiments are then performed, with each fold used in turn as the validation set, and the remaining 9 folds as the training set. Therefore, each data point is used once for testing and 9 times for training, and the result is 10 experimental LogitBoost models trained on 90% fractions. The validation results from 10 experimental models are then combined to provide a measure of the overall performance (Figure 2).¹³

Statistical analysis

Continuous variables are presented as mean \pm SD. The performance of ML to predict ACM was compared with CCTA severity scores (SSS, SIS, and DI) and FRS, using receiver-operating characteristic (ROC) analysis and pairwise comparisons according to Delong *et al.*¹⁸ For ML scores, the optimal thresholds for three risk categories (low, intermediate, and high) were defined using the two-graph-ROC (TG-ROC) technique at the 90% accuracy level.¹⁹ Calibration of the LogitBoost model was assessed using the Brier score method (range: 0–1)—lower scores being consistent with better model calibration.²⁰ Calibration was also assessed graphically with plots showing the observed and predicted proportion of events, grouped by decile of risk.²¹ For all analyses, surviving

patients were right censored to their follow-up date. All statistical tests were two tailed and $P < 0.05$ was considered significant.

Results

Study population

Table 1 describes the baseline characteristics of the study population. There was no drop-out in this analysis and primary outcome data were determined for all patients. Seven hundred and forty-five patients died during the mean follow-up time of 5.4 ± 1.4 years.

Feature selection

Figure 1 shows the results from one representative fold of the 10-fold cross-validation procedure. In this representative fold, using the information gain ranking criteria, 54 of the available 69 variables were selected for the LogitBoost model (19 clinical and 35 from CCTA). As expected, age was the highest ranked feature for ACM prediction, but 8 of the top 10 ranked features were CCTA-derived factors rather than clinical.

Prediction all-cause mortality

Machine learning exhibited a higher area-under-the-curve (AUC) compared with FRS or CCTA data alone for prediction of 5-year ACM (ML: 0.79 vs. FRS: 0.61, SSS: 0.64, SIS: 0.64, DI: 0.62; $P < 0.001$ for all) (Figure 3). Segment stenosis score and SIS were superior to FRS for predicting ACM ($P < 0.05$).

Risk categorization by machine learning

Based on TG-ROC analysis, the intermediate-risk range for the LogitBoost model was a risk probability estimate of between 3.8 and 14%, corresponding to 90% sensitivity and specificity, respectively.

Table 1 Baseline patient characteristics

Characteristic	Data (n = 10 030)
Age (years \pm SD)	59 \pm 13
Sex, n (%)	
Male	5628 (56)
Female	4402 (44)
CAD risk factors, n (%)	
Diabetes	1783 (18)
Hypercholesterolaemia	5415 (54)
Hypertension	5477 (55)
Current smoker	2432 (24)
Family history	3936 (39)
Chest pain, n (%)	
Non-cardiac	1615 (16)
Atypical angina	2967 (30)
Typical angina	1820 (18)
Dyspnoea on exertion, n (%)	3224 (32)

CAD, coronary artery disease.

Low- and high-risk ranges were considered as either side of the intermediate range. With this categorization, the number of patients assigned to low-, intermediate-, and high-risk groups in the study population by the LogitBoost model were: 3960 (39%), 4768 (48%), and 1302 (13%), respectively. The corresponding observed 5-year ACM event rates in these risk-groups were 1.8, 6.7, and 27%, respectively.

Calibration of machine learning scores

The Brier score for the LogitBoost model predicting ACM was 0.08 indicating good calibration between LogitBoost scores (estimated predicted risk) and observed 5-year risk. A calibration plot also confirmed good agreement between LogitBoost scores and the observed 5-year risk of ACM (Figure 4).

Discussion

In this study, we observed ML methods to be an effective method for the prediction of 5-year ACM. The ML method performed superiorly to clinical and CCTA summary metrics alone. To our knowledge, this is the first large-scale evaluation of ML for prognostic risk assessment using CCTA data. The observed efficacy suggests ML has an important clinical role in evaluating prognostic risk in individual patients with suspected CAD.

The ability to correctly identify high-risk patients who may benefit from intensified preventative measures is a major challenge in

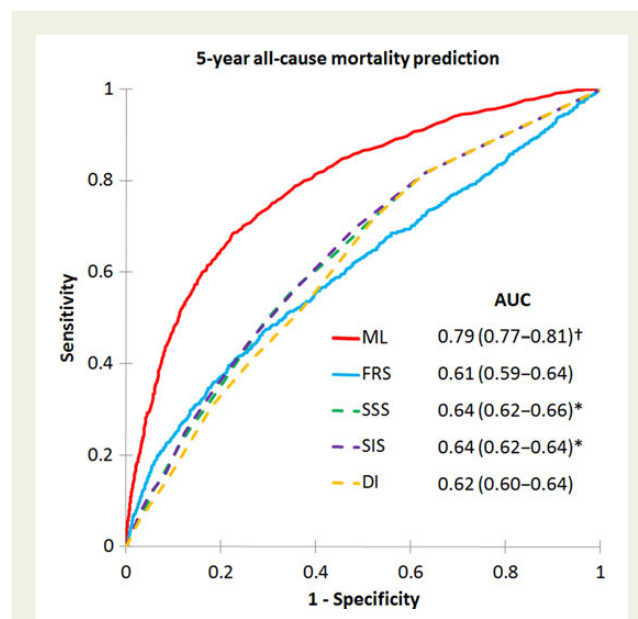


Figure 3 Receiver-operating characteristic curves for prediction of 5-year all-cause mortality. Machine learning using feature selection and a LogitBoost model had a significantly higher area-under-the-curve for all-cause mortality prediction than all other scores ($P < 0.001$)[†]. Area-under-the-curves for segment stenosis score and segment involvement score were greater than Framingham risk score ($P < 0.05$)[‡]. FRS, Framingham risk score; SSS, segment stenosis score; SIS, segment involvement score; DI, modified Duke index.

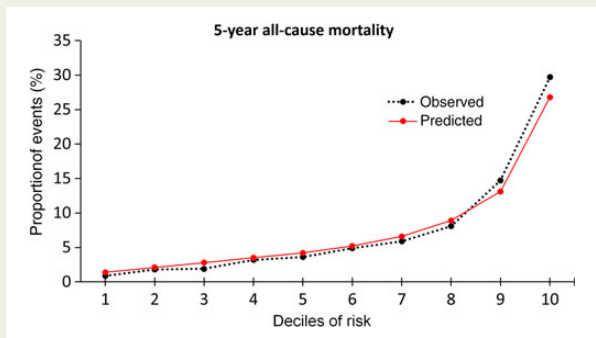


Figure 4 Calibration plot for LogitBoost model. The calibration plot shows the relationship between the observed and predicted proportion of events, grouped by decile of risk. The LogitBoost model showed good calibration with the observed 5-year risk of all-cause mortality.

cardiovascular medicine.^{22,23} Although traditional risk factors for CAD offer general guidance and are effective tools for population-based comparisons, they are ineffective for individual risk assessments.^{23,24} Even when combined into global summary scores for easy usage, the use of clinical variables alone is associated with substantial over- or under-treatment, emphasizing the need for more precise assessment through patient-specific imaging.^{10,23}

Appreciating and integrating the myriad risk predictors in an individual patient is a challenge for the clinician. To date, efforts to improve risk-stratification by using CCTA have largely relied upon luminal stenosis severity. The emphasis placed on this variable over others is in alignment with prior studies using invasive coronary angiography, but ignores an array of other variables important in the CAD pathogenic process, including coronary calcium content, plaque composition, and plaque burden.^{2,25} As an increasing number of CCTA variables along with all clinical variables affecting risk need to be considered, the complexity of assessment increases, making it more difficult for a clinician to draw an overall conclusion regarding risk in an individual patient. Furthermore, the potential influence of unexpected interactions between several weaker predictors in an individual patient is often overlooked. In this study, we demonstrate that ML is able to overcome these challenges, by providing deep integration of the comprehensive CCTA and clinical data.

We performed risk assessment using state-of-the-art ML methods, a form of artificial intelligence that distinguishes itself from traditional prognostic methods by making no *priori* assumptions about causative factors. This characteristic allows for an agnostic exploration of all available data for non-linear patterns that may predict a particular individual's risk, i.e. precision risk-stratification. This important concept represents a divergence from a 'hypothesis-driven' approach conventional in traditional prognostic risk assessments. As observed in the present study, the lack of any required hypothesis avoids overlooking important but unexpected predictor variables or interactions, as well as enabling the recognition of clinically important risk among patients with several marginal risk factors (or no risk factors at all). Further, by its nature, any given ML algorithm requires only minimal input during its model-building phase and none after that. This feature of ML methods is particularly

important, given the ease with which the machine can seamlessly incorporate new data to continually update and optimize its algorithm—and thus continually improve its predictive performance over time.^{5,26}

In a prior study, our group employed a traditional logistic regression approach to predict 2-year ACM, also based on data from the CONFIRM registry.² The resulting model comprised three clinical and CCTA variables, and exhibited an AUC of 0.68. In the current study, which extends follow-up to 5 years, ML methods offered an incremental gain in prognostic performance while handling over 60 variables and a vast number of variable–variable interactions in each patient. The latter effectively individualizes risk assessment and overcomes many of the limitations of a standard statistical approach.

The results of this study have considerable clinical import. Prior large-scale studies have documented the inadequacies of current limited-variable methods for prediction of disease prevalence, severity, and prognosis; and highlight the need for more accurate methods to perform these important clinical tasks.^{27–29} As observed in other fields that require accurate prediction and decision-making in the presence of large amounts of data, ML may offer an improved alternative to current cardiovascular risk-stratification in the individual patient. These improvements may be furthered by the emergence of standard electronic health records and software capable of automated atherosclerotic plaque characterization and quantification—both of which may improve any given ML algorithm.

A limitation of the ML approach is that by using an increasing number of variables and interactions to predict risk, it can subsequently be difficult to identify the specific therapeutic targets that will reduce risk in that individual patient. However, the current therapeutic approach to CAD is very similar across all risk groups, and the benefit of an incremental improvement in risk-stratification in this context, even if a specific causative risk factor or interaction cannot be identified, is that it may allow physicians to move away from this 'one-size fits all' approach with confidence—with the advantage of cost-savings and reduced adverse effects in those that currently receive unnecessary treatments. Furthermore, in time those patients in whom ML re-classifies risk different from traditional approaches can be prospectively studied in greater depth to identify the causative factors and interactions—and this may finally unlock new therapeutic targets.

Study limitations

In the current study, we have only been able to discuss the feasibility and performance of ML, but not its prospective practical implementation. The ability of any ML approach to handle a large number of variables is only advantageous if these variables are actually inputted into the ML algorithm—and this can be a time-consuming process, depending on the mode of data entry. However, we envisage ML working in the background of standard electronic health records or clinical reporting systems, gathering its variables automatically without additional burden on the physician—allowing on-the-fly risk score computation. This is the same principle already adopted by many ML applications—for example, the personalized advertisements that appear in real-time while web-browsing are all based on the passive collection of variables and their seamless input into ML algorithms. In this regard, the feature selection component is particularly important to future practical implementation as it

minimizes the amount of data needed to be handled by such an on-the-fly ML application without compromising on accuracy.

Although prospectively collected, the CONFIRM registry data used to derive the LogitBoost model was nevertheless observational, and the influence of selection bias cannot be discounted. Observational cohort studies, however, are the best suited to assess the initial feasibility of ML algorithms, given the requirement to both develop and validate their efficacy. Once validated, prospective randomized controlled treatment trials based upon ML risk-stratification can then be considered. Whether an objective ML approach without prior hypothesis is more effective than subjective but flexible human thinking is an issue that future prospective studies may explore. Future prospective studies may also allow comparison of our ML approach integrating CCTA and clinical data, to the alternative strategy in this study population of clinical assessment with stress-testing.

Although alternative, more cardiac-specific endpoints than ACM may have been utilized, these data were not available for all centres in the CONFIRM registry. For the current study which had the aim of demonstrating feasibility of the ML approach, ACM was the most definitive end-point, with the maximal amount of patients to facilitate the best trained and validated ML model.

Although we evaluated 69 distinct variables with ML, along with variable–variable interactions, we did not consider the totality of features that may offer improved risk prediction, as we were limited to those collected for the CONFIRM registry. In particular, more detailed clinical data such as duration, magnitude and control of risk factors, as well more advanced plaque characterization with % aggregate plaque volume, arterial remodelling, or Hounsfield-unit-based plaque composition classification, may augment the current ML model.

Another limitation is the use of FRS as the summary metric for evaluation of clinical parameters as it was originally validated for 10-year outcomes in asymptomatic patients. However, this is a commonly adopted strategy given the absence of any more suitable clinical risk score for symptomatic patients or shorter follow-up periods, and moreover, it frames the accuracy of the ML-model in the context of a widely used and understood clinical score.^{2–4} Finally, while numerous ML methods exist, we evaluated the efficacy of only a single ensemble method (LogitBoost). In particular, boosting is also available for Cox-based models (CoxBoost) which leads to the possibility of time-to-event prediction using such alternative analysis.³⁰

Conclusions

In a large, prospective 5-year multinational study of patients with suspected CAD, ML was an effective method for the prediction of ACM, and performed superiorly to the use of clinical and CCTA variables alone.

Authors' contributions

M.M., P.J.S., H.G. performed statistical analysis; J.K.M., D.S.B., P.J.S. handled funding and supervision; M.M., D.D., S.A., M.A.-M., D.A.M.J.B., F.C., T.Q.C., H.-J.C., K.C., B.J.W.C., R.C., A.D., M.G., H.G., M.H., J.H., N.H., G.F., P.A.K., Y.-J.K.; J.L., F.Y.L., E.M., H.M., G.P., G.R., R.R., L.J.S., J.S., T.C.V., A.D., D.S.B., J.K.M., P.J.S. acquired

the data; P.J.S., M.M., D.S.B., J.K.M. conceived and designed the research; M.M., P.J.S. drafted the manuscript; D.D., D.S.B., J.K.M. made critical revision of the manuscript for key intellectual content.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Acknowledgements

We thank all staff at the participating sites for their effort to collect data for the CONFIRM registry.

Funding

This work was supported by research grants from the National Heart Lung and Blood Institute: R01-HL111141, R01-HL115150, R01-HL118019 (J.K.M.), R01-2R01HL089765 (P.J.S.); the Dowager Countess Eleanor Peel Trust, Manchester, UK, and the Dickinson Trust, Manchester, UK (M.M.); National Research Foundation of Korea (Leading Foreign Research Institute Recruitment Program 2012027176) (H.J.C.); Adelson Family Foundation (D.S.B) as well as a generous gift from the Dalio Foundation, New York, NY and the Michael Wolk Foundation, New York, NY (J.K.M.).

Conflict of interest: none declared.

References

- Budoff MJ, Dowe D, Jollis JG, Gitter M, Sutherland J, Halamert E, Scherer M, Bellinger R, Martin A, Benton R, Delago A, Min JK. Diagnostic performance of 64-multidetector row coronary computed tomographic angiography for evaluation of coronary artery stenosis in individuals without known coronary artery disease: results from the prospective multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) trial. *J Am Coll Cardiol* 2008;**52**:1724–1732.
- Hadamitzky M, Achenbach S, Al-Mallah M, Berman D, Budoff M, Cademartiri F, Callister T, Chang HJ, Cheng V, Chinnaiyan K, Chow BJW, Cury R, Delago A, Dunning A, Feuchtner G, Gomez M, Kaufmann P, Kim YJ, Leipsic J, Lin FY, Maffei E, Min JK, Raff G, Shaw LJ, Villines TC, Hausleiter J. Optimized prognostic score for coronary computed tomographic angiography: Results from the CONFIRM registry (CORonary CT angiography evaluation for clinical outcomes: An international multicenter registry). *J Am Coll Cardiol* 2013;**62**:468–476.
- Hadamitzky M, Freibmuth B, Meyer T, Hein F, Kastrati A, Martinoff S, Schömig A, Hausleiter J. Prognostic value of coronary computed tomographic angiography for prediction of cardiac events in patients with suspected coronary artery disease. *JACC Cardiovasc Imaging* 2009;**2**:404–411.
- Chow BJ, Small G, Yam Y, Chen L, Achenbach S, Al-Mallah M, Berman DS, Budoff MJ, Cademartiri F, Callister TQ, Chang H-J, Cheng V, Chinnaiyan K, Delago A, Dunning A, Hadamitzky M, Hausleiter J, Karlsberg RP, Kaufmann P, Lin F, Maffei E, Raff GL, Shaw LJ, Villines TC, Min JK. Incremental prognostic value of cardiac CT in CAD using CONFIRM (CORonary computed tomography angiography evaluation for clinical outcomes: an International Multicenter registry). *Circ Cardiovasc Imaging* 2011;**4**:463–472.
- Waljee AK, Higgins PDR. Machine learning in medicine: a primer for physicians. *Am J Gastroenterol* 2010;**105**:1224–1226.
- Deo RC. Machine learning in medicine. *Circulation*. 2015;**132**:1920–1930.
- Dietterich TG. Ensemble methods in machine learning. *Lect Notes Comput Sci* 2000; **1857**:1–15.
- Min JK, Dunning A, Lin FY, Achenbach S, Al-Mallah MH, Berman DS, Budoff MJ, Cademartiri F, Callister TQ, Chang H-J, Cheng V, Chinnaiyan KM, Chow B, Delago A, Hadamitzky M, Hausleiter J, Karlsberg RP, Kaufmann P, Maffei E, Nasir K, Pencina MJ, Raff GL, Shaw LJ, Villines TC. Rationale and design of the CONFIRM (CORonary CT Angiography Evaluation For Clinical Outcomes: An International Multicenter) Registry. *J Cardiovasc Comput Tomogr* 2011;**5**:84–92.
- Abbara S, Arbab-Zadeh A, Callister TQ, Desai MY, Mamuya W, Thomson L, Weigold WG. SCCT guidelines for performance of coronary computed tomographic angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee. *J Cardiovasc Comput Tomogr* 2009;**3**:190–204.
- Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;**97**:1837–1847.

11. Hall M, Holmes G. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Trans Knowledge Data Eng* 2003;**15**:1437–1447.
12. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann Stat* 2000;**28**:337–407.
13. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;**21**:3301–3307.
14. Arsanjani R, Dey D, Khachatryan T, Shalev A, Hayes S, Fish M, Nakanishi R, Germano G, Berman D, Slomka P. Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population. *J Nucl Cardiol* 2015;**22**:877–884.
15. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD Explor News* 2009;**11**:10–18.
16. Kanamori T, Takenouchi T, Eguchi S, Murata N. Robust loss functions for boosting. *Neural Comput*. 2007;**19**:2183–2244.
17. Witten IH, Frank E, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Boston: Morgan Kaufman; 2011.
18. 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.
19. Greiner M, Sohr D, Göbel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J Immunol Methods* 1995;**185**:123–132.
20. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950;**78**:1–3.
21. Selvarajah S, Kaur G, Haniff J, Cheong KC, Hiong TG, van der Graaf Y, Bots ML. Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population. *Int J Cardiol* 2014;**176**:211–218.
22. Franco M, Cooper RS, Bilal U, Fuster V. Challenges and opportunities for cardiovascular disease prevention. *Am J Med* 2015;**124**:95–102.
23. Arbab-Zadeh A, Fuster V. The myth of the “vulnerable plaque”. *J Am Coll Cardiol* 2015;**65**:846–855.
24. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, Wilson PW. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA* 2003;**290**:891–897.
25. Budoff MJ, Shaw LJ, Liu ST, Weinstein SR, Mosler TP, Tseng PH, Flores FR, Callister TQ, Raggi P, Berman DS. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. *J Am Coll Cardiol* 2007;**49**:1860–1870.
26. Mjolsness E, DeCoste D. Machine learning for science: state of the art and future prospects. *Science* 2001;**293**:2051–2055.
27. Patel MR, Peterson ED, Dai D, Brennan JM, Redberg RF, Anderson HV, Brindis RG, Douglas PS. Low diagnostic yield of elective coronary angiography. *N Engl J Med* 2010;**362**:886–895.
28. Meijboom WB, Mieghem CA, van Pelt N, Weustink A, Pugliese F, Mollet NR, Boersma E, Regar E, van Geuns RJ, de Jaegere PJ, Serruys PW, Krestin GP, de Feyter PJ. Comprehensive assessment of coronary artery stenoses computed tomography coronary angiography versus conventional coronary angiography and correlation with fractional flow reserve in patients with stable angina. *J Am Coll Cardiol* 2008;**52**:636–643.
29. Kang X, Berman DS, Lewin HC, Cohen I, Friedman JD, Germano G, Hachamovitch R, Shaw LJ. Incremental prognostic value of myocardial perfusion single photon emission computed tomography in patients with diabetes mellitus. *Am Heart J* 1999;**138**:1025–1032.
30. Binder H, Allignol A, Schumacher M, Beyersmann J. Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics* 2009;**25**:890–896.