

Risk vs. disease: the role of artificial intelligence in avoiding unneeded testing

Filippo Cademartiri * and Alberto Clemente

Department of Radiology, Fondazione Toscana Gabriele Monasterio (FTGM)-CNR, Pisa, Italy

Online publish-ahead-of-print 12 February 2022

This editorial refers to ‘Preventing unnecessary imaging in patients suspect of coronary artery disease through machine learning of electronic health records’, by L.M. Overmars et al., pp. 11–19.

Introduction

In this issue of the Journal, Overmars et al.¹ published a study entitled ‘Preventing unnecessary imaging in patients suspect of coronary artery disease through machine learning of electronic health records’. They developed algorithms trained on routinely available electronic health records (EHRs), raw electrocardiograms, and blood samples data, to exclude coronary artery disease (CAD) in patients prior to any other clinical or instrumental assessment.¹ They conclude that their algorithm has a very high negative predictive value for the exclusion of CAD (0.96/0.97 based on anatomic modelling and 0.75–0.92 based on functional modelling—i.e. ischaemia imaging) and the expenses of a very low specificity. Their effort in modelling is quite impressive and provides interesting insight into the topic of alternative diagnostic approaches. There are some issues with the study which are fairly and properly mentioned and addressed by the authors; mostly they have to do with the quality and quantity of the data available for the modelling in the first place.

Discussion

A massive amount of scientific resources has been devoted in the past decades towards the creation and improvement of tools for the stratification of individuals and patients with suspected or known CAD. However, over the years, there has always been the practical observation that all models that have been developed and implemented were not delivering the expected performance in every day clinical routine. At the same time, there has been raising awareness that risk stratification is an epidemiological concept and tool, hence, it does not work for diagnostic purposes; this was probably one of the seeds for the concept of personalized medicine. The risk of being sick

is different from actually being sick, even though in a very early and pre-clinical phase of the disease.

In this landscape, there is also a progressively deeper awareness of the costs of healthcare which leads to the search for newer and smarter approaches.

Massive dataset available for each individual or patient (EHRs) may open the possibility for risk stratification but also for early or very early diagnosis. And yet again we should be very clear: risk is not disease. Approaching diagnosis as we approach risk stratification is probably wrong.

Therefore, we should talk about ‘Pre-Diagnosis’, and Pre-Diagnostic(s) as the field in which we study methods for diagnosing a disease in a very early phase (*Figure 1*); and what we normally address as risk factors should be substituted by disease factors.

The availability of a multitude of structured and unstructured EHRs represents an opportunity for the implementation of innovation and offers new chances for the development, monitoring, evaluation, and control of decision-making processes and the implementation of new policy strategies.

In the last decade, the widespread availability of machine learning (ML) and artificial intelligence (AI) tools, associated with increasing computational power, has boosted the attempt to further improve this field of research based on the capability to feed AI algorithms with massive amount of data.

Unfortunately, we still rely very much on the advancement and performance of ‘conventional’ diagnostic tools and algorithms.

The first objective which may have a significant impact in Pre-Diagnostic field may be to identify individuals with no disease and separate them from the rest of the population.

Semi-automatic data management algorithms are therefore desirable to streamline decision-making procedures, reduce subjective evaluation errors and ensure a balance between the cost and benefit of decision-making procedures.

Existing models of obstructive CAD by the European Society of Cardiology (ESC)^{2–4} or the American Heart Association (AHA)⁵ postulate a logistic regression model of relatively few traditional

The opinions expressed in this article are not necessarily those of the Editors of the *European Heart Journal – Digital Health* or of the European Society of Cardiology.

*Corresponding author. Tel: +39 050 3152216, Email: filippocademartiri@gmail.com

© The Author(s) 2022. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

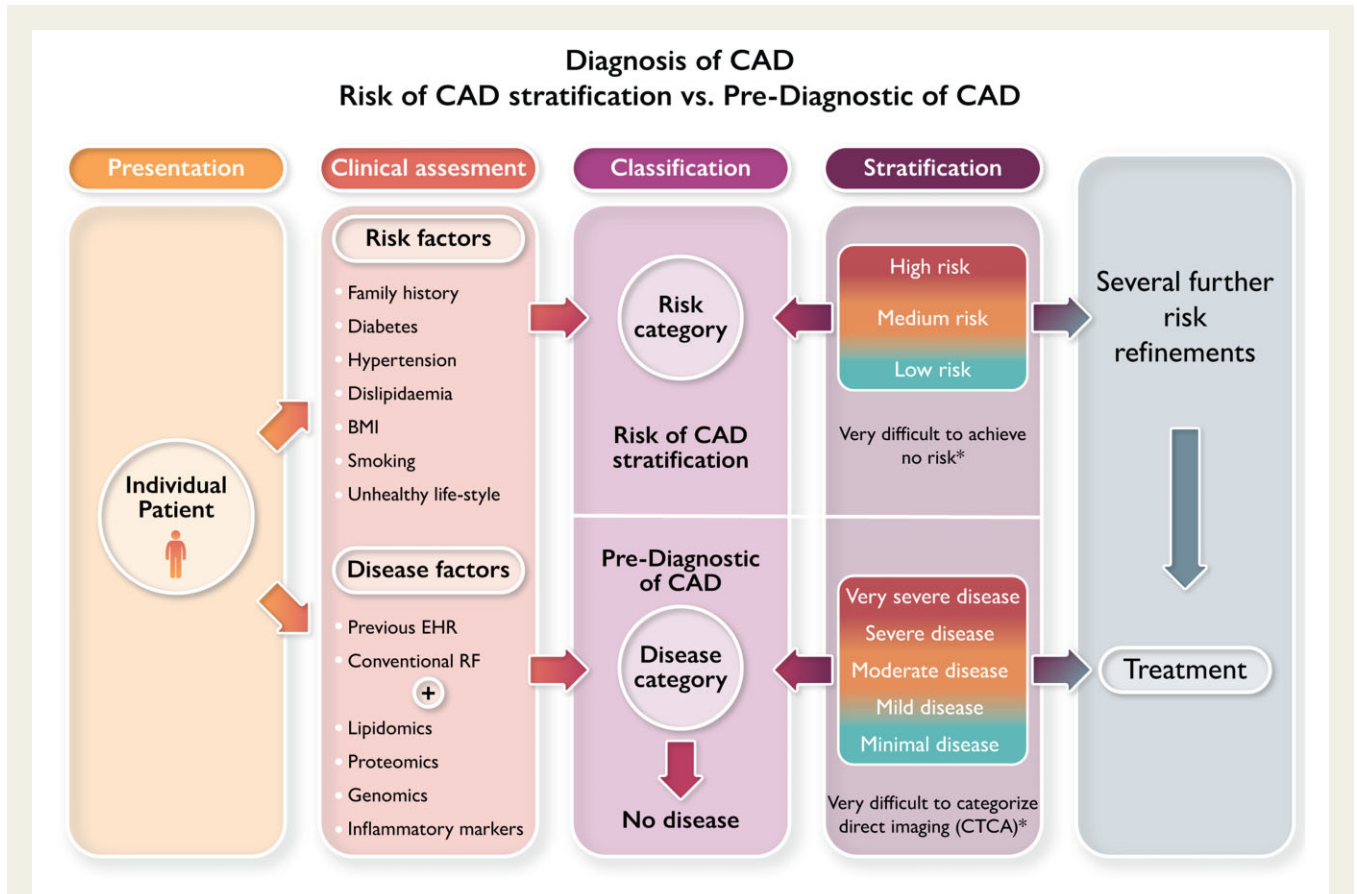


Figure 1 The figure shows the two different concepts of risk of coronary artery disease stratification (contemporary approach) and of Pre-Diagnostic of coronary artery disease (forthcoming approach). Our current approach is based on constant modulation and re-modulation of risk of coronary artery disease (in this case not of major adverse cardiac/cardiovascular events) until the treatment phase is achieved and this corresponds to the concept of risk in the figure. The newer approach that may be allowed by adequate artificial intelligence-based models is the Pre-Diagnostic of coronary artery disease with early exclusion of all individuals with no coronary artery disease and direct referral to coronary computed tomography angiography; this is more consistent with a disease oriented medicine. AI, artificial intelligence; CAD, coronary artery disease; CTCA, coronary computed tomography angiography; EHR, electronic health record; RF, risk factor; RM, regression models.

disease predictors. Despite the reported good performance of such parametric regression models, a systematic review demonstrated poor external validation and head-to-head comparisons, poor reporting of their technical characteristics as well as variability in outcome variables, predictors and prediction horizons, which limits their applicability in evidence-based decision-making in healthcare.⁶

Moreover, the increasing availability of large data sets and the highly improved computational power, seem to have directed large part of recent research towards model development rather than model validation⁷; in other words, we have several models with little or no validation.

Machine learning and AI could be used for this purpose by enabling the identification of the most informative features from big data, which now are becoming available, incorporating several features, ranging from clinical examinations and lab tests to advanced analytics such as lipidomics, proteomics, and genomics.⁸

Furthermore, recent computational models could be used as prognostic tools or as treatment tools in the case of implementing computational biomechanics models of virtual stenting applications.^{9,10}

Although advanced imaging equipment (e.g. computed tomography, magnetic resonance, nuclear medicine equipments, and so forth) manufacturers provide analysis software able to collect several qualitative/quantitative information, to the best of our knowledge, there is no clinically validated platform available on the market that has a clinical decision support system (CDSS) which integrates imaging-based and non-imaging-based models.

Besides the computationally based CDSS systems, cyber-physical systems in the form of point of care devices have been developed to easily and cost-effectively measure biomarkers, which can be used for the diagnosis of CAD. The bio-nanochip system measures several biomarkers (for example in saliva and blood), offering diagnostic accuracy equal to laboratory methods.¹¹

To our knowledge, the only attempt to structure a complex CDSS is H2020-SMARTool project (GA number: 689068).¹² In addition, plasma lipidomics may be a promising source of diagnostic and prognostic biomarkers in cardiovascular disease, exploitable not only to assess the risk of adverse events but also to

identify subjects without coronary atherosclerosis, thus reducing unnecessary testing.¹³

In conclusion, there are great efforts and expectations in the field of ML/AI implementation on EHR; however, there are also significant issues in the management of quality and quantity of data which is key to algorithm training and performance in clinical routine.

Conflict of interest: none declared.

References

- 1). Overmars LM, van Es B, Groepenhoff F, et al. Preventing unnecessary imaging in patients suspect of coronary artery disease through machine learning of electronic health records. *Eur Heart J Digit Health* 2021;**3**:11–19.
- 2). Knuuti J, Ballo H, Juarez-Orozco LE, et al. The performance of non-invasive tests to rule-in and rule-out significant coronary artery stenosis in patients with stable angina: a meta-analysis focused on post-test disease probability. *Eur Heart J* 2018;**39**:3322–3330.
- 3). Knuuti J, Wijns W, Saraste A, et al.; ESC Scientific Document Group. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;**41**:407–477. Erratum in: *Eur Heart J* 2020 Nov 21; **41**(44):4242.
- 4). Bing R, Singh T, Dweck MR, et al. Validation of European Society of Cardiology pre-test probabilities for obstructive coronary artery disease in suspected stable angina. *Eur Heart J Qual Care Clin Outcomes* 2020;**6**:293–300.
- 5). Hendel RC. Pretest probability: cornerstone of testing in suspected ischemic heart disease: a call to revise criteria for noninvasive testing. *Circ Cardiovasc Imaging* 2019;**12**:e009835.
- 6). Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;**353**:i2416.
- 7). Mincarone P, Bodini A, Tumolo MR, et al. Discrimination capability of pretest probability of stable coronary artery disease: a systematic review and meta-analysis suggesting how to improve validation procedures. *BMJ Open* 2021;**11**:e047677.
- 8). Elashoff MR, Wingrove JA, Beineke P, et al. Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. *BMC Med Genomics* 2011;**4**:26.
- 9). Sakellarios A, Bourantas CV, Papadopoulou SL, et al. Prediction of atherosclerotic disease progression using LDL transport modelling: a serial computed tomographic coronary angiographic study. *Eur Heart J Cardiovasc Imaging* 2017;**18**:11–18.
- 10). Sakellarios AI, Tsompou P, Kigka V, et al. Non-invasive prediction of site-specific coronary atherosclerotic plaque progression using lipidomics, blood flow, and LDL transport modeling. *Appl Sci* 2021;**11**:1976.
- 11). Christodoulides N, Pierre FN, Sanchez X, et al. Programmable bio-nanochip technology for the diagnosis of cardiovascular disease at the point-of-care. *Methodist Debaquey Cardiovasc J* 2012;**8**:6–12. Erratum in: *Methodist Debaquey Cardiovasc J* 2012;**8**(3):48.
- 12). Sakellarios AI, Rigas G, Kigka V, et al. SMARTool: a tool for clinical decision support for the management of patients with coronary artery disease based on modeling of atherosclerotic plaque process. *Annu Int Conf IEEE Eng Med Biol Soc* 2017;**2017**:96–99.
- 13). Bodini A, Michelucci E, Di Giorgi N, et al. Predictive added value of selected plasma lipids to a re-estimated minimal risk tool. *Front Cardiovasc Med* 2021;**8**:682785.