

Improving risk stratification in patients with coronary artery disease using artificial intelligence

Mitchel Arno Molenaar

Colophon

The research for this dissertation was conducted at the Amsterdam University Medical Center, Department of Cardiology, Amsterdam, the Netherlands. It was made possible by an innovation grant from the Amsterdam University Medical Center, in close collaboration and support of Image Guided Therapy Systems – Philips, Best, the Netherlands.

ISBN: 978-94-6522-761-0

Cover design: Lyonne van Gaalen

Printed by: Ridderprint | www.ridderprint.nl

© 2025 Mitchel Arno Molenaar

All rights reserved. No part of this thesis may be reproduced, stored, or transmitted in any form or by any means without prior written permission of the author or the copyright-owning journals for published chapters.

Improving risk stratification in patients with coronary artery disease using
artificial intelligence

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 20 november 2025, te 16.00 uur

door Mitchel Arno Molenaar

geboren te Schagen

Promotiecommissie

Promotor: prof. dr. S.A.J. Chamuleau AMC-UvA

Copromotores: dr. M.J. Schuurings
dr. N.J. Verouden Universiteit Twente
AMC-UvA

Overige leden: prof. dr. A. Abu-Hanna AMC-UvA
prof. dr. M.P. Schijven AMC-UvA
prof. dr. F.M.A.C. Martens Vrije Universiteit Amsterdam
prof. dr. R. Nijveldt Radboud Universiteit
dr. J.M. Wolterink Universiteit Twente
prof. dr. H.A. Marquering AMC-UvA

Faculteit der Geneeskunde

Contents

Chapter 1 General introduction	9
Chapter 2 The impact of valvular heart disease in patients with chronic coronary syndrome	19
Chapter 3 Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome	39
Chapter 4 Validation of machine learning-based risk stratification scores for patients with acute coronary syndrome treated with percutaneous coronary intervention	65
Chapter 5 Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease	85
Chapter 6 Deep learning-based segmentation of coronary arteries and stenosis detection in x-ray coronary angiography	103
Chapter 7 General discussion	133
Summary	148
Samenvatting (Dutch summary)	151
List of publications	154
Portfolio	156
Acknowledgments	158
Curriculum Vitae/About the author	162

1

General introduction

This introduction aims to explain the key aspects of coronary artery disease (CAD) and current treatment strategies, as well as the role of risk stratification in clinical decision making and the value of artificial intelligence (AI) in enhancing risk prediction. This information will lay the foundation for the research presented in the subsequent chapters of this thesis.

Coronary artery disease

CAD is one of the most common diseases characterized by atherosclerotic plaque buildup in the coronary arteries, which leads to their narrowing and reduced blood flow to the heart muscle (Figure 1)^(1, 2). Worldwide there are approximately 126 million individuals affected by this disease⁽³⁾, leading to approximately 9 million deaths each year⁽⁴⁾.

The development of CAD is influenced by multiple factors, including genetic, environmental, and lifestyle factors. The presentation of CAD is heterogeneous and can be divided into two categories: chronic coronary syndrome (CCS) and acute coronary syndrome (ACS)^(1, 2). Each has its distinct causes, symptoms, and management^(1, 5). ACS is characterized by a plaque rupture that results in thrombus formation, leading to decreased blood flow to the myocardium⁽⁶⁾. The manifestation of ACS is acute, with chest discomfort as most common symptom in both men and women⁽⁷⁾. CCS tends to develop progressively, caused by the gradual buildup of plaques. The ischemic symptoms in CCS are caused by the mismatch between the myocardial oxygen supply and demand, which is commonly induced by physical activity or emotional stress⁽¹⁾.

Treatment options

To reduce symptoms, slow the progression of atherosclerosis, and prevent cardiovascular events, CCS is typically managed with lifestyle changes and medical therapy. In some patients, revascularization by percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG) is indicated to reduce symptoms and/or improve prognosis. Patients suspected of ACS, with or without clinical signs of ST segment elevation on electrocardiography, receive antithrombotic therapy and are considered for invasive coronary angiography (ICA) in the catheterization laboratory (cath lab). In addition to the anatomical information provided by ICA regarding the thrombus burden and atherosclerotic lesions, functional (pressure) measurements can offer additional information about the impact on coronary blood flow^(2, 5). In high risk patients, ACS requires immediate restoration of the blood flow to the myocardium, performed by balloon angioplasty and/or PCI⁽⁵⁾. Patients with obstructive and clinically important stenoses in the coronary arteries are discussed in a multidisciplinary heart team that decides on the revascularization strategy (PCI/CABG) and timing of revascularization. Therapeutic decisions are guided by risk stratification.

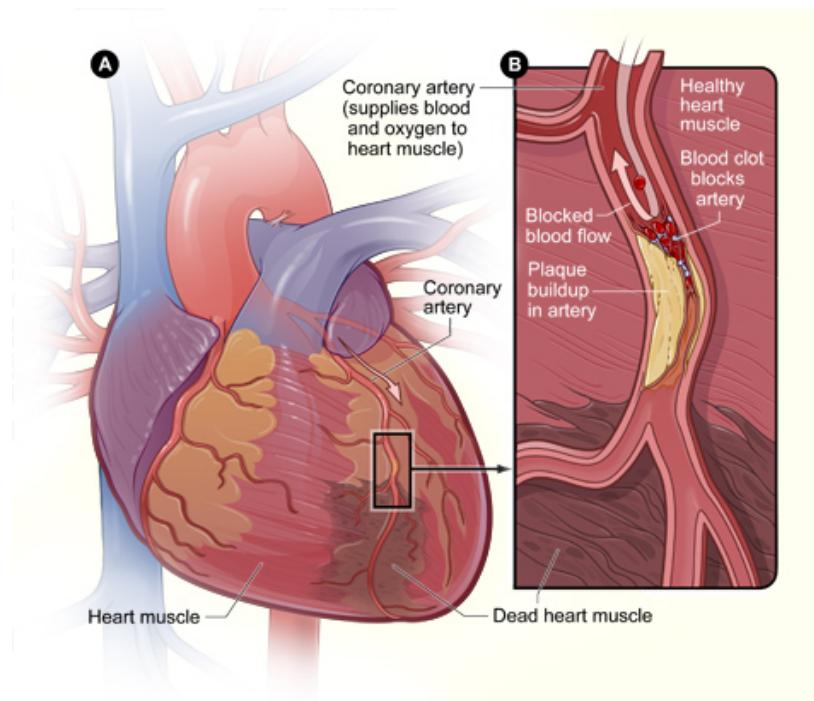


Figure 1: Coronary artery disease is caused by plaque buildup, leading to narrowing of the coronary arteries. Rupture of these plaques results in a blood clot that blocks the blood flow, which leads to oxygen insufficiency and damage to the heart muscle. Image "Heart With Muscle Damage and a Blocked Artery" used with permission by National Heart, Lung and Blood Institute, National Institutes of Health, licensed under CC BY-NC 2.0. Source: <https://www.flickr.com/photos/nihgov/33328945325>.

Risk stratification

Risk stratification is the process of categorizing patients based on their risk on adverse outcomes, as determined through risk prediction. Risk stratification has several advantages. It helps identify patients at high or rising risk who may benefit from more intensive care, prioritize interventions, and allocate healthcare resources. There are several risk stratification tools available in patients with CAD, which are summarized below.

Risk stratification in chronic coronary syndrome

In patients with CCS, risk stratification consists of clinical assessment, non-invasive assessment of ischemia or coronary anatomy, and, in minority of the cases, invasive coronary angiography (ICA)⁽²⁾. The European Society of Cardiology (ESC) 2024 Guidelines on CCS recommend transthoracic echocardiographic (TTE) measurement of left ventricular (LV) function for risk stratification in patients with CCS, as it is considered to be one of the strongest determinants of mortality^(1, 2). Patients with a reduced LV function are at high risk with a cardiac death of >3% per year^(1, 2). This is mainly based on the Coronary Artery Surgery Study (CASS) registry⁽⁸⁾, a registry that was

performed in the 1970s, a long time before introduction of modern therapies and revascularization techniques. However, the patient's condition encompasses more than LV function alone. To further guide clinicians in the process of risk stratification, risk scores/calculators are developed using demographic information, medical history, risk factors and laboratory measurements. Examples of risk scores for the prediction of cardiovascular events in patients with CCS are the Framingham risk score⁽⁹⁾, ABC-CHD model⁽¹⁰⁾, and Duke Treadmill score⁽¹¹⁾. Other risk scores that are not specifically developed for patients with CCS are the SCORE2⁽¹²⁾/SCORE-OP⁽¹³⁾ and SMART risk score⁽¹⁴⁾. These risk scores have limitations. First, most risk scores are developed using linear models, which may not capture non-linear or complex relationships between the variables and the predicted outcome. Second, these risk scores consist of a limited set of traditional risk factors (e.g. age, blood pressure, smoking, diabetes, and cholesterol), which do not incorporate other non-cardiovascular risk factors (e.g. malignancy) or other relevant patient information, such as genetic or imaging information. Third, current risk stratification tools are focused on assigning the individual patient to a specific risk category (e.g. high, mid or low risk). Within these risk categories, it is difficult to identify the patients who will truly experience the event and who will benefit from an intervention^(15, 16).

Risk stratification in acute coronary syndrome

In ACS patients, multiple risk scores have been developed, including the Global Registry of Acute Coronary Events (GRACE) risk score^(5, 17, 18), Thrombolysis in Myocardial Infarction (TIMI)^(19, 20), Prediction of Adverse Events Following an Acute Coronary Syndrome (PRAISE)⁽²¹⁾, and The Platelet glycoprotein IIb/IIIa in Unstable angina: Receptor Suppression Using Integrilin Therapy (PURSUIT)⁽²²⁾. Of these risk scores, the GRACE version 2.0 risk score has shown the best discriminative performance. Its calculation is therefore recommended by the ESC guidelines for risk stratification in patients with ACS^(5, 17, 18). The GRACE 2.0 score offers an estimate of the risk of in-hospital and six-month mortality using eight clinical variables. Patients with a high GRACE 2.0 score (>140) should be considered for early angiography (within 24 hour)⁽⁵⁾. Although the GRACE 2.0 risk score has been validated by multiple studies⁽²³⁾, recent studies showed a suboptimal calibration (agreement between predicted and actual outcomes)^(24–26), and a reduced discriminative performance (the degree to which a model can differentiate between patients who are likely to die and those who are not) in women⁽²⁵⁾.

Severity of coronary artery disease

A risk score that helps clinicians decide in the choice between the revascularization strategies, PCI or CABG, is the Synergy Between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery (SYNTAX) score. The latest introduced SYNTAX II score uses both coronary anatomy data, to assess the burden of CAD, and clinical prognostic data. However, scoring the anatomic elements of the SYNTAX score from two-dimensional ICA images is difficult and prone to subjective

assessments, which leads to intra- and inter-observer variability^(27–30). In particular, bifurcation lesions are prone to inconsistent assessment⁽²⁷⁾.

1

Artificial intelligence for clinical assistance

AI refers to the broad concept of computer systems performing tasks that previously required human intelligence. It is believed that Alan Turing (1912–1954), a British mathematician, laid the foundation for a machine's ability to learn from experience by publishing his article “Computing Machinery and Intelligence” in 1950⁽³¹⁾. From that moment, AI has undergone significant advancements, particularly over the past three decades, due to the advances of algorithms, computational power, and availability of large datasets. In the last decade, the number of publications on machine learning applications in healthcare has grown exponentially. Machine learning is a subfield of AI and can be interpreted as a set of techniques to enable AI. Machine learning involves the use of models to learn from data and make predictions⁽³²⁾. In cardiology, these applications have been developed to support clinicians in the assessment of cardiac structures and function⁽³³⁾, detection of diseases or abnormalities⁽³⁴⁾, and prediction of outcomes⁽³⁵⁾, using clinical, imaging⁽³⁶⁾, and signal data⁽³⁷⁾.

Machine learning using echocardiography data

Machine learning has been applied to automate parts of the clinical workflow of TTE with the aim to reduce time and workload, and improve reproducibility, standardization and clinical outcome⁽³⁸⁾. For example, machine learning models have been developed for automated image view classification⁽³⁹⁾, automated measurements (estimating ejection fraction, a measure of LV function)⁽³⁴⁾ and disease detection⁽⁴⁰⁾. Although the models showed good performance, only a limited number were evaluated in an external cohort to assess their generalizability⁽³⁸⁾. Previous studies have shown promising results for the prediction of adverse outcomes using TTE video or image data⁽⁴¹⁾ or data extracted from TTE images^(42, 43). Despite the importance of TTE for risk stratification in patients with CCS, the value of machine learning for CCS patients using TTE data remains unexplored and has been limited to stress echocardiography⁽⁴⁴⁾.

Deep learning using invasive coronary angiography data

In ICA, prior studies mainly focused on automated analysis of these images with the aim to assist clinicians in its interpretation that is both reproducible and objective. Deep learning models, a subtype of machine learning based on artificial neural networks, have been developed to automate the selection of key frames⁽⁴⁵⁾, identify coronary arteries (segmentation)^(46–48), and detect and classify lesions^(49, 50). However, most image interpretation models were trained on single-center datasets and validation in an external center is often lacking. The same applies to models to predict clinical outcomes in patients with CAD⁽⁵¹⁾.

Aim and outline of the thesis

In summary, CAD is a common multifactorial disease in which accurate risk prediction is challenging. Currently available risk stratification tools for patients with CAD are constrained by a limited number of included personalized features, have limitations in performance, or are subject to intra- and/or inter-observer variability. Machine learning may enhance risk stratification in patients with CAD by improving risk prediction using data from TTE and ICA, both of which play an essential role in the diagnosis, risk stratification, and treatment of CAD. Despite promising results of machine learning-based risk scores, there are limited data in patients with CAD, and validation is often lacking.

The aim of this thesis is to investigate the value of machine learning for risk stratification of patients with CAD, with a specific focus on TTE and ICA data. To achieve this aim, this thesis is divided into two parts and concludes with a general discussion.

Part 1 – Non-invasive imaging in patients with coronary artery disease

Main research question: Can machine learning-based models improve the risk stratification of patients with CCS utilizing both TTE and clinical data?

LV function is an established predictor of mortality in patients with CCS. It is unclear whether other echocardiographic parameters provide additional information about the prognosis of patients with CCS. Therefore, in **chapter 2**, we investigated impact of valvular heart disease on mortality in patients with CCS in a retrospective cohort study.

Machine learning may improve risk prediction in CCS patients, thereby supporting more accurate risk stratification. In **chapter 3**, we evaluated the accuracy of machine learning using clinical and TTE data to predict all-cause five-year mortality in patients with CCS and compared its performance to traditional risk scores.

Part 2 – Invasive imaging in patients with coronary artery disease

Main research questions: Can machine learning-based models improve the risk stratification of patients with ACS treated with PCI? Can machine learning models utilizing ICA imaging improve risk stratification?

Recently, machine learning-based risk scores have been introduced and made widely available. The validation of these models on data from an external center is crucial to investigate their generalizability. In **chapter 4**, we validated the machine learning-based GRACE version 3.0 risk score and PRAISE risk score in patients with ACS treated with PCI for predicting mortality at a Dutch tertiary center. Both risk scores were compared to the GRACE version 2.0 risk score.

In literature, there was no clear overview of the current applications of machine learning for automated imaging analysis of ICA. **Chapter 5** summarizes the current applications of AI for automated imaging analysis of ICA, supported by reviewed literature. The fundamental concepts of AI are further explained, and the potential future applications in the cath lab are described.

Chapter 6 presents the performance of deep learning-based segmentation and lesion detection models in ICA data, which were trained and validated on large datasets from two tertiary centers.

General discussion

1

Chapter 7 summarizes the main findings presented in this thesis and discusses its impact on clinical care and future research. In addition, the current extent to which machine learning-based risk stratification has been implemented in clinical practice for patients with CAD will be discussed.

In conclusion, this thesis presents research on the value of machine learning for risk stratification of patients with CAD, with a specific focus on TTE and ICA data.

References

- [1] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;41:407–477.
- [2] Vrints C, Andreotti F, Koskinas KC, Rossello X, Adamo M, Ainslie J, et al. 2024 ESC Guidelines for the management of chronic coronary syndromes. *Eur Heart J*. 2024;45:3415–3537.
- [3] Dai H, Much AA, Maor E, Asher E, Younis A, Xu Y, et al. Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990–2017: results from the Global Burden of Disease Study 2017. *European Heart Journal - Quality of Care and Clinical Outcomes* 2022;8:50–60.
- [4] Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 2018;392:1736–1788.
- [5] Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, et al. 2023 ESC Guidelines for the management of acute coronary syndromes: Developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *European Heart Journal* 2023;44:3720–3826.
- [6] Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of Plaque Formation and Rupture. *Circulation Research* 2014;114:1852–1866.
- [7] Oosterhout REM van, Boer AR de, Maas AHEM, Rutten FH, Bots ML, Peters SAE. Sex Differences in Symptom Presentation in Acute Coronary Syndromes: A Systematic Review and Meta-analysis. *Journal of the American Heart Association* 2020;9:e014733.
- [8] Emond M, Mock MB, Davis KB, Fisher LD, Holmes DR, Chaitman BR, et al. Long-term survival of medically treated patients in the Coronary Artery Surgery Study (CASS) Registry. *Circulation* 1994;90:2645–2657.
- [9] D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–753.
- [10] Lindholm D, Lindbäck J, Armstrong PW, Budaj A, Cannon CP, Granger CB, et al. Biomarker-Based Risk Model to Predict Cardiovascular Mortality in Patients With Stable Coronary Disease. *J Am Coll Cardiol* 2017;70:813–826.
- [11] Mark DB, Shaw L, Harrell FE, Hlatky MA, Lee KL, Bengtson JR, et al. Prognostic Value of a Treadmill Exercise Score in Outpatients with Suspected Coronary Artery Disease. *New England Journal of Medicine* 1991;325:849–853.
- [12] SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal* 2021;42:2439–2454.
- [13] SCORE2-OP working group and ESC Cardiovascular risk collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *European Heart Journal* 2021;42:2455–2467.
- [14] Derosteijn JAN, Visseren FLJ, Wassink AMJ, Gondrie MJA, Steyerberg EW, Ridker PM, et al. Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score. *Heart* 2013;99:866–872.
- [15] Tunstall-Pedoe H. Cardiovascular Risk and Risk Scores: ASSIGN, Framingham, QRISK and others: how to choose. *Heart* 2011;97:442–444.
- [16] De Backer G. New insights in cardiovascular risk estimation and stratification. *E-Journal of Cardiology Practice* 2022;22. Available from: <https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-22/new-insights-in-cardiovascular-risk-estimation-and-stratification> [Accessed 8 January 2024].
- [17] Fox KAA, Fitzgerald G, Puymirat E, Huang W, Carruthers K, Simon T, et al. Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ Open* 2014;4:e004425.
- [18] D'Ascenzo F, Biondi-Zoccali G, Moretti C, Bollati M, Omede P, Sciuto F, et al. TIMI, GRACE and alternative risk scores in Acute Coronary Syndromes: a meta-analysis of 40 derivation studies on 216,552 patients and of 42 validation studies on 31,625 patients. *Contemp Clin Trials* 2012;33:507–14.
- [19] Antman EM, Cohen M, Bernink PJ, McCabe CH, Horacek T, Papuchis G, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA* 2000;284:835–842.
- [20] Morrow DA, Antman EM, Charlesworth A, Cairns R, Murphy SA, Lemos JA de, et al. TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation: An intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation* 2000;102:2031–2037.
- [21] D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. PRAISE study group. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *Lancet*. 2021;397:199–207.
- [22] Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy (PURSUIT) Trial Investigators. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. *N Engl J Med*. 1998 Aug 13;339(7):436–43. doi: 10.1056/NEJM199808133390704. PMID: 9705684.
- [23] Fox KA, A, Eagle KA, Gore JM, Steg PG, Anderson FA, et al. The Global Registry of Acute Coronary Events, 1999 to 2009—GRACE. *Heart* 2010;96:1095–1101.
- [24] Sangen NMR van der, Azzahhafi J, Yin DRPPCP, Peper J, Rayhi S, Walhout RJ, et al. External validation of the GRACE risk score and the risk-treatment paradox in patients with acute coronary syndrome. *Open Heart* 2022;9:e001984.
- [25] Wenzl FA, Krämer S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score

- in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *The Lancet* 2022;400:744–756.
- [26] Moledina SM, Kontopantelis E, Wijeyasundera HC, Banerjee S, Van Spall HGC, Gale CP, et al. Ethnicity-dependent performance of the Global Registry of Acute Coronary Events risk score for prediction of non-ST-segment elevation myocardial infarction in-hospital mortality: nationwide cohort study. *Eur Heart J* 2022;43:2289–2299.
- [27] Garg S, Girasis C, Sarno G, Goedhart D, Morel M-A, Garcia-Garcia HM, et al. The SYNTAX score revisited: a reassessment of the SYNTAX score reproducibility. *Catheter Cardiovasc Interv* 2010;75:946–952.
- [28] Généreux P, Palmerini T, Caixeta A, Cristea E, Mehran R, Sanchez R, et al. SYNTAX score reproducibility and variability between interventional cardiologists, core laboratory technicians, and quantitative coronary measurements. *Circ Cardiovasc Interv* 2011;4:553–561.
- [29] Serruys PW, Chichareon P, Modolo R, Leaman DM, Reiber JHC, Emanuelsson H, et al. The SYNTAX score on its way out or ... towards artificial intelligence: part II. *EurolIntervention* 2020;16:60–75.
- [30] Basman C, Levine E, Tejpal A, Thampi S, Rashid U, Barry R, et al. Variability and Reproducibility of the SYNTAX Score for Triple-Vessel Disease. *Cardiovascular Revascularization Medicine* 2022;37:86–89.
- [31] Turing, AM. I.—Computing Machinery and Intelligence. *Mind* 1950;LIX:433–460.
- [32] Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial Intelligence in Cardiology. *J Am Coll Cardiol*. 2018;71:2668–2679.
- [33] Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. *npj Digit Med* 2020;3:10.
- [34] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;580:252–256.
- [35] Sabouri M, Rajabi AB, Hajianfar G, Gharibi O, Mohebi M, Avval AH, et al. Machine learning based readmission and mortality prediction in heart failure patients. *Sci Rep* 2023;13:18671.
- [36] Siegersma KR, Leiner T, Chew DP, Appelman Y, Hofstra L, Verjans JW. Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. *Neth Heart J* 2019;27:403–413.
- [37] Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med* 2020;26:886–891.
- [38] Schuurings MJ, İlsgüm I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;8:648877.
- [39] Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 2018;1:6.
- [40] Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit Med*. 2018;1:59.
- [41] Ulloa Cerna AE, Jing L, Good CW, vanMaanen DP, Raghunath S, Suever JD, et al. Deep-learning-assisted analysis of echocardiographic videos improves predictions of all-cause mortality. *Nat Biomed Eng* 2021;5:546–554.
- [42] Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, et al. Predicting Survival From Large Echocardiography and Electronic Health Record Datasets: Optimization With Machine Learning. *JACC Cardiovasc Imaging* 2019;12:681–689.
- [43] Kwon J, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography* 2019;36:213–218.
- [44] Gaibazzi N, Cortigiani L, Ciampi Q, Lorenzoni G, Rigo F, Gherardi S, et al. Machine-learning algorithms for prediction of survival by stress echocardiography in chronic coronary syndromes. *European Heart Journal* 2022;43:e hac544.079.
- [45] Ciusdel C, Turcea A, Puin A, Itu L, Calmac L, Weiss E, et al. Deep neural networks for ECG-free cardiac phase and end-diastolic frame detection on coronary angiographies. *Computerized Medical Imaging and Graphics* 2020;84:101749.
- [46] Cervantes-Sanchez F, Cruz-Aceves I, Hernandez-Aguirre A, Hernandez-Gonzalez MA, Solorio-Meza SE. Automatic Segmentation of Coronary Arteries in X-ray Angiograms using Multiscale Analysis and Artificial Neural Networks. *Applied Sciences* 2019;9:5507.
- [47] Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep* 2019;9:1–11.
- [48] Du T, Xie L, Zhang H, Liu X, Wang X, Chen D, et al. Training and validation of a deep learning architecture for the automatic analysis of coronary angiography. *EurolIntervention* 2021;17:32–40.
- [49] Zhao C, Vij A, Malhotra S, Tang J, Tang H, Pienta D, et al. Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Computers in Biology and Medicine* 2021;136:104667.
- [50] Danilov VV, Klyshnikov KY, Gerget OM, Kutikhin AG, Ganyukov VI, Frangi AF, et al. Real-time coronary artery stenosis detection based on modern neural networks. *Sci Rep* 2021;11:7582.
- [51] Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External Validations of Cardiovascular Clinical Prediction Models: A Large-Scale Review of the Literature. *Circ Cardiovasc Qual Outcomes* 2021;14:e007858.

2

The impact of valvular heart disease in patients with chronic coronary syndrome

Mitchel A. Molenaar

Berto J. Bouma

Casper F. Coerkamp

Jelle P. Man

Ivana Išgum

Niels J. Verouden

Jasper L. Selder

Steven A.J. Chamuleau

Mark J. Schuuring

Abstract

Background

The European Society of Cardiology 2019 Guidelines on chronic coronary syndrome (CCS) recommend echocardiographic measurement of the left ventricular function for risk stratification in all patients with CCS. Whereas CCS and valvular heart disease (VHD) share common pathophysiological pathways and risk factors, data on the impact of VHD in CCS patients are scarce.

Methods

Clinical data including treatment and mortality of patients diagnosed with CCS who underwent comprehensive transthoracic echocardiography (TTE) in two tertiary centers were collected. The outcome was all-cause mortality. Data were analyzed with Kaplan-Meier curves and Cox proportional hazard analysis adjusting for significant covariables and time-dependent treatment.

Results

Between 2014 and 2021 a total of 1,984 patients with CCS (59% men) with a median age of 65 years (interquartile range [IQR] 57–73) underwent comprehensive TTE. Severe VHD was present in 44 patients and moderate VHD in 325 patients. A total of 654 patients (33%) were treated with revascularization, 39 patients (2%) received valve repair or replacement and 299 patients (15%) died during the median follow-up time of 3.5 years (IQR 1.7–5.6). Moderate or severe VHD (hazard ratio=1.33; 95% CI 1.02–1.72) was significantly associated with mortality risk, independent of LV function and other covariables, as compared to no/mild VHD.

Conclusions

VHD has a significant impact on mortality in patients with CCS additional to LV dysfunction, which emphasizes the need for a comprehensive echocardiographic assessment in these patients.

Introduction

Chronic coronary syndrome (CCS) is characterized by stable atherosclerotic coronary plaques that build up over time. CCS affects more than 34 million adult Europeans and has a high mortality rate, despite advanced medical care and revascularization^(1, 2).

The European Society of Cardiology (ESC) 2019 Guidelines on CCS recommend transthoracic echocardiographic (TTE) measurement of left ventricular (LV) function for risk stratification in all patients with CCS, as it is considered to be one of the strongest determinants of mortality^(3–5). This is mainly based on the Coronary Artery Surgery Study (CASS) registry⁽⁵⁾. This CASS registry was performed in the 1970s, a long time before introduction of modern therapies and revascularization techniques.

It is unclear whether other echocardiographic findings provide additional information about the prognosis of patients with CCS. In particular data regarding the impact of valvular heart disease (VHD) on the mortality of patients with CCS are scarce. It is estimated that VHD affects more than 18 million Europeans^(6, 7) and accounts for 10 to 20% of all cardiac surgery procedures^(8, 9). As both VHD and CCS share common pathophysiological pathways and risk factors, a better understanding of the impact of VHD on mortality in patients with CCS is needed. Therefore, we performed a study in CCS patients and investigated the impact of VHD on mortality.

Methods

Design and patient population

In this retrospective cohort study, patients diagnosed by the treating physician with CCS between 2014 and 2021 were consecutively selected from electronic health records of the Amsterdam University Medical Center (two tertiary centers), the Netherlands. CCS patients (≥ 18 years) who underwent comprehensive transthoracic echocardiography (TTE) one year before or after the outpatient visit were eligible for inclusion. For patients with multiple studies, the TTE closest to the outpatient visit date was selected. This retrospective cohort study was approved by the local institutional review board, who waived the need for written consent.

Data collection and echocardiographic measurements

Baseline, treatment and mortality in follow-up data of patients with CCS were collected from pseudonymized electronic health records and stored in a registry. Two-dimensional TTE with Doppler tissue imaging (Vivid 9, GE Vingmed Ultrasound AS, Horten, Norway; Philips Epiq, Philips Affiniti and Philips IE33, Philips Medical Systems, Best, The Netherlands) was performed and assessed by clinical technicians or cardiology residents according to recommendations of the European Association of Cardiovascular Imaging⁽¹⁰⁾, ESC guidelines^(11, 12) and standard operating procedure⁽¹³⁾. TTE images were digitized and analyzed using vendor-specific software (GE EchoPAC, GE Vingmed Ultrasound AS, Horten, Norway; Xcelera, Philips Medical Systems, Best, The Netherlands; TomTec 2D Cardiac Performance Analysis, Munich, Germany).

The initial assessment of the valves was performed qualitatively by a clinical technician or cardiology resident. Semi-quantitative and quantitative measurements of stenosis or regurgitation were obtained if indicated, especially if clinical decisions were based on these findings^(11, 12). The results were documented in a TTE report⁽¹⁰⁾, which were overseen by dedicated imaging cardiologists who made corrections if needed to maintain accuracy and completeness.

The Simpson's method of disks was used to estimate the LV volume at both end-diastolic and end-systolic phase from apical four- and two-chamber views. This method involved tracing the endocardial border on 2D echocardiographic images of the left ventricle during these phases, dividing the tracing into multiple disks (slices), and summing their volumes to obtain the total LV volume. The LV ejection fraction (LVEF) was calculated by subtracting the LV volume at end-systole from the LV volume at end-diastole, and dividing it by the LV volume at end-diastole. LV dysfunction was defined as mild to severe abnormal LV function (LVEF 51% for male and 53% for female). Moderately and severely abnormal LV function were defined as a LVEF of 30%–41% and <30%, respectively⁽¹⁴⁾.

For the purpose of this study, the LV function and severity of aortic stenosis (AS), aortic regurgitation (AR), mitral stenosis (MS), mitral regurgitation (MR), tricuspid regurgitation (TR), pulmonary stenosis (PS), and pulmonary regurgitation (PR) were stored in the registry. Patients were excluded if TTE image acquisition was of poor quality or incomplete due to missing assessment of LV function or VHD⁽¹⁰⁾. Multivalvular disease was defined as regurgitation and/or stenosis in two or more heart valves. Patients were categorized based on the most severe valve condition among the valves, which means that patients with both moderate and severe valvular lesions were classified as having severe VHD. Impaired renal function was defined as an estimated glomerular filtration rate (eGFR)<60 mL/min/1.73 m². eGFR was calculated with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation⁽¹⁵⁾. Obesity was defined as a body mass index of >30 kg/m².

Outcome

The clinical endpoint was all-cause mortality. The end of follow-up was defined as the last recorded contact with the tertiary center or the date of mortality.

Statistical analysis

Results were expressed as mean values with standard deviation (SD) for normally distributed data, and median with interquartile range (IQR) for not normally distributed data. Nominal or ordinal data were expressed with numbers and percentages. The Shapiro-Wilk test was used to test for normality. The one-way ANOVA or Kruskal-Wallis test was performed for between group comparisons of continuous data, as appropriate. A Pearson's Chi-Square test was performed for categorical variables.

Missing values were imputed by multiple imputation by chained equation (MICE) with a linear regression model, which was iteratively performed for 10 iterations. The degree of multicollinearity between variables was assessed with the variance inflation factor (VIF)⁽¹⁶⁾. Variables

with a VIF greater than 10 were either dichotomized, centered by subtracting the mean value, or omitted from multivariable analysis to account for their collinearity with other variables. Variables with a p-value of $p \leq 0.05$ on univariable mortality analysis and time-dependent variables coronary revascularization and valve repair or replacement were entered into Cox proportional hazards (PH) models with backward selection procedure ($p \leq 0.05$) based on the Akaike information criterion. Cox PH analyses were conducted to examine the association between predictor variables and mortality. Predictor variables with a prevalence of at least 5% among the patients were included in multivariable analysis. To evaluate the assumption of PH, Kaplan Meier curves were inspected and Schoenfeld residuals were calculated⁽¹⁷⁾. Analysis was performed for the predictor variables VHD, the number of valves affected and the specific subtypes of VHD (AS, AR, MS, MR, TR, PS and PR).

Statistical analyses were performed in Python V3.8 and RStudio V.2022.07.0 (RStudio Team, Boston, MA) using R-version 4.1.3. (R Core Team, Vienna, Austria). $P < 0.05$ was considered statistically significant.

Results

Study population

A total of 2,845 patients with CCS were screened. Among them, 861 patients were not eligible for inclusion, resulting in a study population of 1,984 patients, as shown in Supplementary Figure 1. The proportion of missing data was 3%, as shown in Supplementary Table 1. The median age of the study population was 65 years (IQR 57–73), 59% were men and 26% had a history of myocardial infarction (Supplementary Table 2). The majority, 54%, had a history of hypertension and 59% of the patients presented with chest pain at the outpatient visit. A minority of the patients presented with dyspnea (31%). The most common reported secondary prevention therapies were antiplatelet therapy (61%), statins (60%), and beta-blockers (52%). A total of 505 patients (25%) had LV dysfunction, which was severe in 39 patients (2%), as shown in Supplementary Table 3.

Concomitant valvular heart disease

No/mild, moderate and severe VHD were present in 1,615 (82%), 325 (16%) and 44 (2%) patients, respectively (Supplementary Table 3). MR was most the common VHD (moderate MR: 176 patients, severe MR: 3 patients) in patients with CCS, followed by TR (moderate TR: 128 patients, severe TR: 16 patients), and AS (moderate AS: 64 patients, severe AS: 24 patients). Multivalvular disease was present in 84 patients (4%). Compared to patients within the group with no/mild VHD, patients with moderate or severe VHD were significantly older and had more often hypertension, atrial fibrillation or flutter, and chronic obstructive pulmonary disease (COPD), used more often cardiovascular medication (anticoagulants, renin-angiotensin system inhibitors, beta-blockers, and diuretics), had a lower eGFR, and more often LV dysfunction (Supplementary

Tables 2 and 3). Patients with moderate or severe VHD had more often surgical aortic valve replacement (SAVR) in medical history. Patients with no/mild VHD were more often a current or former smoker and had more often a family history of coronary artery disease. Chest pain was more often reported in patients with no/mild VHD, while dyspnea was more frequently reported in patients with moderate or severe VHD.

Table 1: Follow-up of study population

	All patients (n=1984)	No/mild VHD (n=1615)	Moderate VHD (n=325)	Severe VHD (n=44)	p-value
Revascularization ^a , n (%)	654 (33.0)	517 (32.0)	120 (36.9)	17 (38.6)	0.165
CABG, n (%)	56 (2.8)	41 (2.5)	10 (3.1)	5 (11.4)	0.002
PCI, n (%)	622 (31.4)	492 (30.5)	116 (35.7)	14 (31.8)	0.179
Isolated or concomitant valve repair or replacement ^b , n (%)	39 (2.0)	1 (0.1)	14 (4.3)	24 (54.5)	<0.001
SAVR, n (%)	10 (0.5)	0 (0.0)	4 (1.2)	6 (13.6)	<0.001
TAVR, n (%)	26 (1.3)	1 (0.1)	8 (2.5)	17 (38.6)	<0.001
SMVR, n (%)	1 (0.1)	0 (0.0)	0 (0.0)	1 (2.3)	<0.001
TEER, n (%)	3 (0.2)	0 (0.0)	2 (0.6)	1 (2.3)	<0.001
STVR, n (%)	1 (0.1)	0 (0.0)	0 (0.0)	1 (2.3)	<0.001
SPVR, n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	NA
Outcome, n (%)					
Death	299 (15.1)	195 (12.1)	91 (28.0)	13 (29.5)	<0.001

^aSome patients received both coronary artery bypass graft and percutaneous coronary intervention.

^bOne patient with mild and 14 patients with moderate valvular heart disease received valve repair or replacement during follow-up due to multivalvular disease, increase in severity, or combined with coronary artery bypass grafting.

CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention; SAVR, surgical aortic valve replacement; SMVR, surgical mitral valve replacement; SPVR, surgical pulmonary valve replacement; STVR, surgical tricuspid valve replacement; TAVR, transcatheter aortic valve replacement; TEER, transcatheter edge-to-edge repair; VHD, valvular heart disease.

Follow-up

The median follow-up time of patients was 3.5 years (IQR 1.7–5.6). During the follow-up period, a total of 654 patients (33%) received revascularization by coronary artery bypass grafting (CABG) and/or percutaneous coronary intervention (PCI). Follow-up data are shown in Table 1. Patients with severe VHD received revascularization by CABG (11%) more often compared to patients with moderate (3%) or no/mild VHD (3%) during follow-up. Valve repair or replacement was performed in 39 patients (2%) during follow-up. Transcatheter aortic valve replacement (TAVR) was most often performed (26 patients) followed by SAVR (10 patients). One patient with mild VHD and 14 patients with moderate VHD received valve repair or replacement during follow-up due to multivalvular disease, increase in severity, or combined with CABG.

A total of 299 patients (15%) died, of which 91 patients had moderate VHD and 13 patients had severe VHD. The mortality curves of patients with moderate and severe VHD showed significant overlap and had both higher mortality rates compared to those with no/mild VHD (Supplementary Figure 2). Based on the overlap in mortality curves and low number of patients with severe VHD (n=44), it was decided to combine moderate and severe in further analysis (Figure 1).

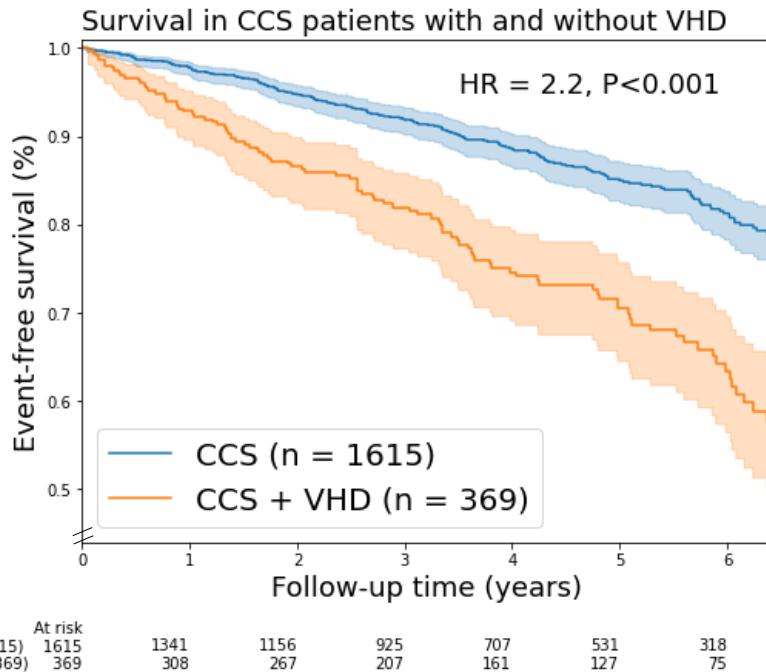


Figure 1: Kaplan-Meier mortality curves for chronic coronary syndrome (CCS) patients with/without valvular heart disease (VHD). In this figure, VHD is moderate and severe combined VHD. Shaded areas represent the 95% confidence intervals. HR, unadjusted hazard ratio.

Determinants of mortality

In univariable analysis, a number of variables were associated with mortality, as shown in Table 2. In multivariable analysis, moderate and severe combined VHD [hazard ratio (HR)=1.33; 95% CI 1.02–1.72] remained associated with mortality independent of age (HR=1.04; 95% CI 1.02–1.05), diabetes (HR=1.58; 95% CI 1.24–2.00), current or former smoking (HR=1.48; 95% CI 1.16–1.88), valve repair or replacement (HR=1.50; 95% CI 1.02–2.20), COPD (HR=1.96; 95% CI 1.42–2.70), chest pain (HR=0.68; 95% CI 0.54–0.85), impaired renal function (HR=1.61; 95% CI 1.27–2.04), and LV dysfunction (HR=1.96; 95% CI 1.55–2.48). Details are shown in Table 2 and Supplementary Figure 3.

Moderate VHD was associated with increased mortality (HR=1.4; 95% CI 1.05–1.78), while severe VHD did not show a significant association with mortality (HR=0.99; 95% CI 0.52–1.87, Supplementary Table 4).

Mortality and multivalvular disease

Patients with multivalvular disease (unadjusted HR=3.2; 95% CI 2.4–4.9) had a higher mortality rate than patients with single valve heart disease (unadjusted HR=1.9; 95% CI 1.5–2.5) and the

group with no/mild VHD (Log-Rank p<0.005, Supplementary Figure 4).

Mortality by VHD subtype

Moderate and severe combined AS (unadjusted HR=1.8, 95% CI 1.2–2.7), AR (unadjusted HR=1.9, 95% CI 1.1–3.1), MR (unadjusted HR=2.3, 95% CI 1.7–3.1), TR (unadjusted HR=2.6, 95% CI 1.9–3.5) were associated with mortality (Figure 2). In addition, moderate and severe combined MS (n=6), PR (n=7) and PS (n=1) were associated with mortality. In multivariable analysis, only moderate TR remained a significant predictor of mortality (HR=1.6; 95% CI 1.2–2.3, Supplementary Table 5).

Table 2: Univariable and multivariable Cox regression analysis of variables associated with mortality: moderate and severe combined

Variable	Univariable analysis		Multivariable analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Age per year	1.1 (1.1-1.1)	<0.001	1.04 (1.02-1.05)	<0.001
Male	1.3 (1.1-1.7)	0.025		
Hypertension	1.4 (1.1-1.8)	0.0041		
Diabetes	1.5 (1.2-2)	<0.001	1.58 (1.24-2.00)	<0.001
Dyslipidaemia	0.98 (0.77-1.2)	0.85		
Current or former smoker	1.4 (1.1-1.7)	0.0083	1.48 (1.16-1.88)	0.001
Family history of CAD	0.63 (0.48-0.83)	0.0011		
Myocardial infarction	1.5 (1.2-1.9)	<0.001		
PCI	1.2 (0.95-1.5)	0.13		
CABG	1.9 (1.4-2.5)	<0.001		
Valve repair or replacement	2.4 (1.7-3.5)	<0.001	1.50 (1.02-2.20)	0.04
Atrial fibrillation/-flutter	2 (1.5-2.6)	<0.001		
Stroke	1.6 (1-2.4)	0.032		
COPD	3 (2.2-4.1)	<0.001	1.96 (1.42-2.70)	<0.001
Chest pain	0.59 (0.47-0.75)	<0.001	0.68 (0.54-0.85)	0.001
Dyspnea	1.3 (1-1.7)	0.019		
Obesity	1.0 (0.8-1.3)	0.76		
Impaired renal function	2.5 (2.0-3.1)	<0.001	1.61 (1.27-2.04)	<0.001
Cholesterol per mmol/l	0.95 (0.85-1.1)	0.31		
LDL per mmol/l	0.86 (0.76-0.97)	0.018		
Triglyceride per mmol/l	1.1 (1-1.2)	0.031		
Left ventricular dysfunction	2.7 (2.1-3.4)	<0.001	1.96 (1.55-2.48)	<0.001
Moderate or severe VHD	2.2 (1.8-2.8)	<0.001	1.33 (1.02-1.72)	0.03

Univariable and multivariable Cox analysis. Variables with a p-value of p≤0.05 on univariable mortality analysis were entered into the multivariable cox analysis with backward selection procedure. Moderate or severe VHD was a significant determinant of mortality independent of left ventricular dysfunction and other covariables.

CABG, coronary artery bypass grafting; CI, confidence interval; COPD, chronic obstructive pulmonary disease; HR, hazard ratio; LDL, low-density lipoprotein cholesterol; PCI, percutaneous coronary intervention; VHD, valvular heart disease.

Discussion

This study in patients with CCS showed that moderate or severe combined VHD was associated with mortality, independent of LV function and other covariables. In the 1,984 patients studied, patients with CCS and moderate or severe VHD had a higher risk of mortality compared

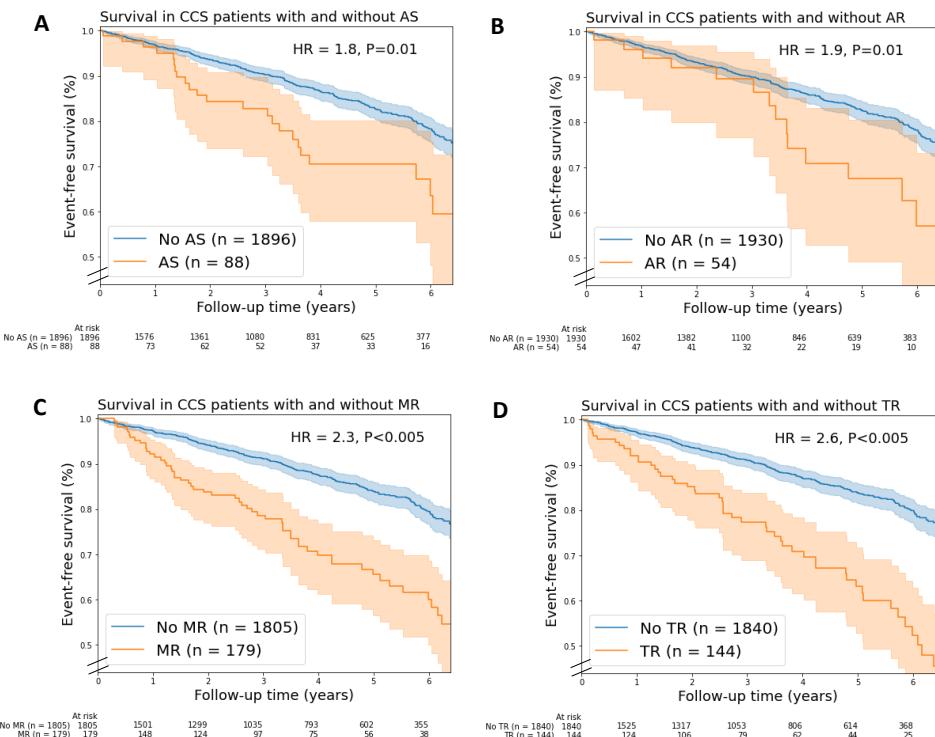


Figure 2: Kaplan-Meier mortality curves for chronic coronary syndrome (CCS) patients with/without (A) aortic stenosis, (B) aortic regurgitation, (C) mitral regurgitation, and (D) tricuspid regurgitation. In this figure, valvular heart disease (VHD) is moderate and severe combined VHD. Shaded areas represent the 95% confidence intervals. AR, aortic regurgitation; AS, aortic stenosis; HR, unadjusted hazard ratio; MR, mitral regurgitation; TR, tricuspid regurgitation.

to patients with CCS only, which increased with the number of valves affected. These findings demonstrate the importance of echocardiographic assessment of VHD, in addition to LV function, in CCS patients.

Valvular heart disease in patients with chronic coronary syndrome

To our knowledge, this is the first study that investigates the prognostic value of VHD in patients with CCS. Severe VHD is an established determinant of mortality for which valve repair or replacement is recommended by current guidelines^(18, 19). We observed that severe VHD was not an independent determinant of mortality in patients with CCS. This finding may be explained by the number of patients with severe VHD (n=44) that may have been too small to detect a significant effect. Moreover, a substantial proportion of CCS patients with severe VHD received a valvular intervention during follow-up (55%) which likely reduced their mortality risk.

We found that moderate VHD was independently associated with mortality. Potential benefit of intervention on mortality for patients with moderate VHD is currently subject of debate⁽²⁰⁾. Ongoing studies are investigating the impact of both moderate AS and LV dysfunction⁽²¹⁾. The

hypothesis that TAVR improves outcomes in these patients is currently being prospectively investigated in the TAVR UNLOAD trial⁽²²⁾. Further prospective investigations are warranted to confirm the prognostic value of VHD in patients with moderate CCS and evaluate the impact of early intervention of VHD on mortality⁽²³⁾. Our findings support the need for improvement of care in VHD patients, which might be achieved by early valve repair or replacement. However, the complexity of the interplay between CCS and VHD on symptoms⁽²⁴⁾, cardiac damage^(25, 26), and clinical course⁽²⁷⁾ make appropriate timing of intervention difficult⁽²⁸⁾.

Previous studies have investigated the association between moderate to severe TR and mortality. In 85%–90% of the patients, TR is caused secondary by left-sided heart failure⁽¹⁸⁾. We found that moderate TR was independently associated with mortality in patients with CCS. Several studies have confirmed that TR is a predictor of mortality independent of LV dysfunction, pulmonary pressures, and right ventricle dilatation and dysfunction^(29–31). Our findings suggest that TR is a marker of advanced disease in patients with CCS, which has more value than merely reflecting the severity of right/left ventricular dysfunction and pulmonary hypertension⁽²⁹⁾.

In this study in two tertiary centers, a significantly higher number of CCS patients had moderate or severe AS (n=88), TR (n=144), and MR (n=179) compared to the general population with the same or older age^(32–34). These relatively high VHD rates were anticipated in these specialized care centers, in which more complex medical conditions are seen. Lower rates of VHD in CCS patients may be expected in non-tertiary centers.

Pathophysiology

Our study shows that VHD has an incremental prognostic role in patients with CCS, which may have several explanations. Firstly, our findings demonstrate that patients with both CCS and moderate or severe VHD have more often risk factors for mortality, including older age, COPD, atrial fibrillation/-flutter, and a lower eGFR. Secondly, patients with both CCS and VHD may have more advanced cardiovascular calcifications, which are observed in atherosclerotic plaque formation, mitral annular calcification, and aortic artery calcification⁽⁴⁾. These calcifications are strong determinants of cardiovascular events⁽³⁵⁾. Thirdly, both CCS and VHD can cause LV dysfunction through ischemia and LV remodeling, which may accelerate deterioration of the LV function leading to end-stage heart failure⁽⁷⁾. Further longitudinal research is necessary to investigate the pathophysiological mechanisms of VHD in patients with CCS.

Clinical implications

The observed independent prognostic value of VHD suggests that it could have a crucial role in the non-invasive risk stratification of patients with CCS. However, despite the advantages of TTE, such as low-costs, portability and absence of radiation⁽³⁶⁾, TTE may not be performed in all patients with CCS as recommended in current guidelines. A recent study by Neglia et al.⁽³⁷⁾ showed that the diagnostic process was not according to the ESC guideline in 44% of the patients with CCS. This finding may have detrimental implications for patients with (suspected) CCS since undiagnosed and untreated VHD is associated with heart failure and mortality⁽³⁴⁾. Therefore,

echocardiography should be performed in all patients with (suspected) CCS to rule out VHD and other cardiac diseases^(4, 38).

The LV function is currently the only recommended echocardiographic assessed feature for risk stratification in patients with CCS. The results of this study indicate that comprehensive echocardiographic assessment of VHD should be included in the standard clinical workup of patients with CCS, at least in those with a normal LV function.

Study limitations

Several remarks can be made about this study. Firstly, the study had a retrospective cohort study design that has inherent limitations. Secondly, to minimize the amount of missing data, information was extracted from textual notes in electronic medical records. The cause of mortality was not available for all patients and was therefore not further differentiated. Thirdly, valve calcifications and quantitative parameters of echocardiography were not evaluated in this study. Fourthly, there may have been a selection bias in this study as the patients who underwent TTE may have had a higher a-priori risk of VHD, which could have influenced the results. Nevertheless, the included patients reflect a real-life population that were seen at a tertiary center.

Conclusions

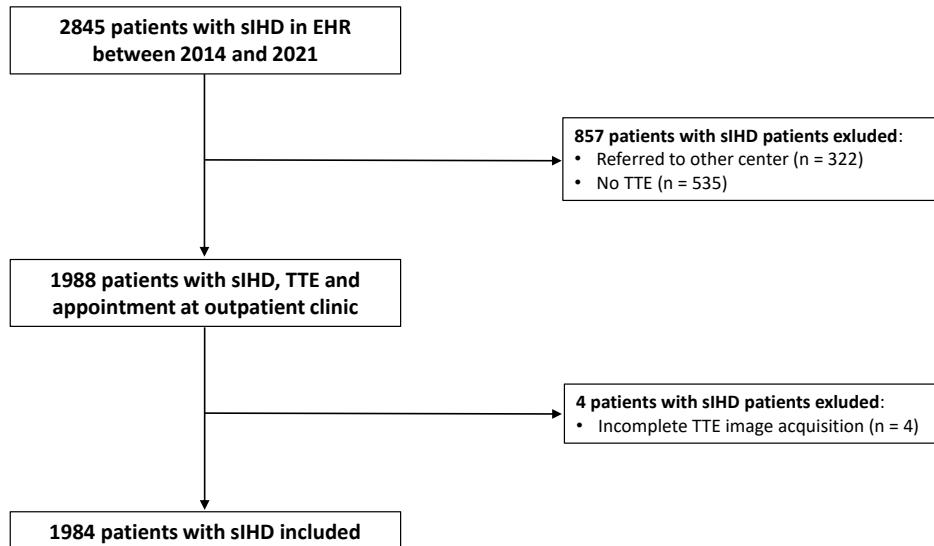
VHD was an independent determinant of mortality in patients with CCS. This finding demonstrates the need for a comprehensive echocardiographic assessment of VHD, in addition to LV function, in CCS patients. Moreover, it indicates that complete assessment of VHD should be included in the standard clinical workup of patients with CCS to improve risk stratification.

References

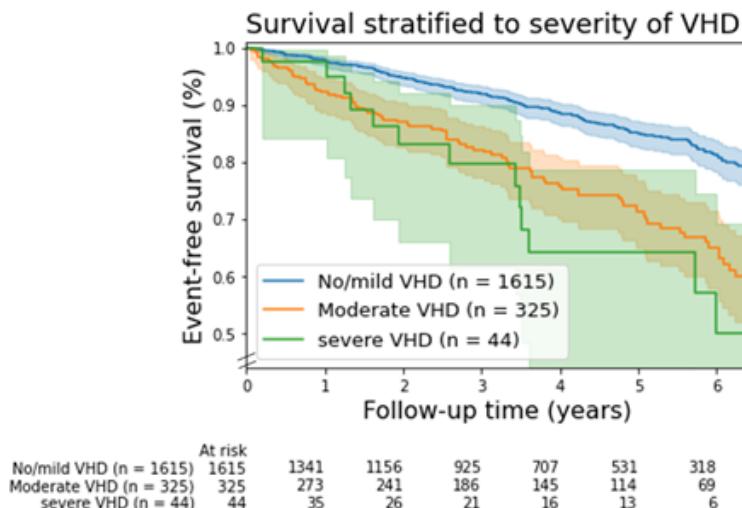
- [1] Sorbets E, Fox KM, Elbez Y, Danchin N, Dorian P, Ferrari R, et al. Long-term outcomes of chronic coronary syndrome worldwide: insights from the international CLARIFY registry. *Eur Heart J* 2020;41:347–56.
- [2] Timmis A, Townsend N, Gale CP, Torbica A, Lettino M, Petersen SE, et al. European Society of cardiology: cardiovascular disease statistics 2019. *Eur Heart J* 2020;41:12–85.
- [3] Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease. *J Am Coll Cardiol* 2012;60:e44–e164.
- [4] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;41:407–77.
- [5] Emond M, Mock MB, Davis KB, Fisher LD, Holmes DR, Chaitman BR, et al. Long-term survival of medically treated patients in the coronary artery surgery study (CASS) registry. *Circulation* 1994;90:2645–57.
- [6] Benjamin EJ, Paul Muntner C, Chair Alvaro Alonso V, Marcio Bittencourt FS, Clifton Callaway MW, April Carson FP. Heart disease and stroke statistics—2019 update: a report from the American heart association. *Circulation* 2019;139:56–28.
- [7] Coisne A, Scotti A, Latib A, Montaigne D, Ho EC, Ludwig S, et al. Impact of moderate aortic stenosis on long-term clinical outcomes: a systematic review and meta-analysis. *JACC Cardiovasc Interv* 2022;15:1664–74.
- [8] Nkomo VT, Gardin JM, Skelton TN, Gottsdiner JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet* 2006;368:1005–11.
- [9] Maganti K, Rigolin VH, Sarano ME, Bonow RO. Valvular heart disease: diagnosis and management. *Mayo Clin Proc* 2010;85:483–500.
- [10] Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European association of cardiovascular imaging. *Eur Heart J* 2017;18:1301–10.
- [11] Vahanian A, Alfieri O, Andreotti F, Antunes MJ, Barón-Esquias G, Baumgartner H, et al. Guidelines on the management of valvular heart disease (version 2012). *Eur Heart J* 2012;33:2451–96.
- [12] Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, et al. 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J* 2017;38:2739–91.
- [13] Bouma BJ, Riezenbos R, Voogel AJ, Veldhorst MH, Jaarsma W, Hrudova J, et al. Appropriate use criteria for echocardiography in The Netherlands. *Neth Heart J* 2017;25:330–4.
- [14] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American society of echocardiography and the European association of cardiovascular imaging. *J Am Soc Echocardiogr* 2015;28:1–39.e14.
- [15] Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–12.
- [16] Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol* 2019;72:558–69.
- [17] Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in cox regression. *Stat Med* 1995;14:1707–23.
- [18] Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, et al. 2021 ESC/EACTS guidelines for the management of valvular heart disease: developed by the task force for the management of valvular heart disease of the European society of cardiology (ESC) and the European association for cardio-thoracic surgery (EACTS). *Eur Heart J* 2022;43:561–632.
- [19] Otto CM, Nishimura RA, Bonow RO, Carabello BA, Erwin JP, Gentile F, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary: a report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. *Circulation* 2021;143:e35–71.
- [20] Stassen J, Galloo X, van der Bijl P, Bax JJ. Focus on diagnosis and prognosis to guide timing of intervention in valvular heart disease. *Curr Cardiol Rep* 2022;24:1407–16.
- [21] Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, et al. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. *JACC Cardiovasc Imaging* 2019;12:681–9.
- [22] Spitzer E, Van Mieghem NM, Pibarot P, Hahn RT, Kodali S, Maurer MS, et al. Rationale and design of the transcatheter aortic valve replacement to UNLOAD the left ventricle in patients with ADvanced heart failure (TAVR UNLOAD) trial. *Am Heart J* 2016;182:80–8.
- [23] Khan KR, Khan OA, Chen C, Liu Y, Kandanelly RR, Jamiel PJ, et al. Impact of moderate aortic stenosis in patients with heart failure with reduced ejection fraction. *J Am Coll Cardiol* 2023;81:1235–44.
- [24] Tastet L, Tribouilloy C, Maréchaux S, Vollema EM, Delgado V, Salaun E, et al. Staging cardiac damage in patients with asymptomatic aortic valve stenosis. *J Am Coll Cardiol* 2019;74:550–63.
- [25] Bohbot Y, Renard C, Manrique A, Levy F, Maréchaux S, Gerber BL, et al. Usefulness of cardiac magnetic resonance imaging in aortic stenosis. *Circulation* 2020;13:e010356.
- [26] Jansen R, Hart EA, Peters M, Urgel K, Kluin J, Tietge WJ, et al. An easy-to-use scoring index to determine severity of mitral regurgitation by 2D echocardiography in clinical practice. *Echocardiography* 2017;34:1275–83.
- [27] Padang R, Bagnall RD, Semsarian C. Genetic basis of familial valvular heart disease. *Circ: Cardiovasc Genet* 2012;5:569–80.
- [28] Lancellotti P, Vannan MA. Timing of intervention in aortic stenosis. *N Engl J Med* 2020;382:191–3.
- [29] Benfari G, Antoine C, Miller WL, Thapa P, Topilsky Y, Rossi A, et al. Excess mortality associated with functional tricuspid regurgitation complicating heart failure with reduced ejection fraction. *Circulation* 2019;140:196–206.

- [30] Nath J, Foster E, Heidenreich PA. Impact of tricuspid regurgitation on long-term survival. *J Am Coll Cardiol* 2004;43:405–9.
- [31] Wang N, Fulcher J, Abeysuriya N, McGrady M, Wilcox I, Celermajer D, et al. Tricuspid regurgitation is associated with increased mortality independent of pulmonary pressures and right heart failure: a systematic review and meta-analysis. *Eur Heart J* 2019;40:476–84.
- [32] Cahill TJ, Prothero A, Wilson J, Kennedy A, Brubert J, Masters M, et al. Community prevalence, mechanisms and outcome of mitral or tricuspid regurgitation. *Heart* 2021;107:1003–9.
- [33] d'Arcy JL, Coffey S, Loudon MA, Kennedy A, Pearson-Stuttard J, Birks J, et al. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the OxVALVE population cohort study. *Eur Heart J* 2016;37:3515–22.
- [34] Dziadzko V, Clavel M-A, Dziadzko M, Medina-Inojosa JR, Michelena H, Maalouf J, et al. Outcome and undertreatment of mitral regurgitation: a community cohort study. *Lancet* 2018;391:960–9.
- [35] Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med* 2008;358:1336–45.
- [36] Schuuring MJ, Işgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;8:648877.
- [37] Neglia D, Liga R, Gimelli A, Podlesnikar T, Cvijić M, Pontone G, et al. Use of cardiac imaging in chronic coronary syndromes: the EURECA imaging registry. *Eur Heart J* 2022;44(2):142–58.
- [38] Schuuring MJ, Tanis W. A remarkable exercise test leading to the diagnosis of left atrial myxoma. *Int J Cardiol* 2015;201:53–4.

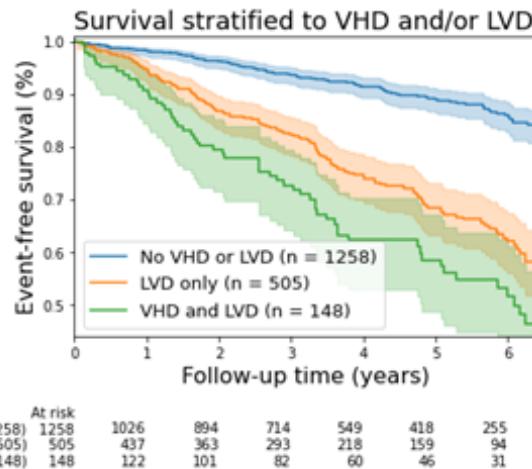
Supplementary material



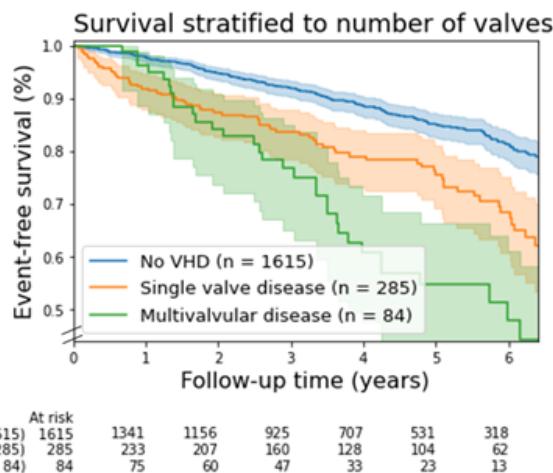
Supplementary Figure 1: Inclusion flowchart. CCS, chronic coronary syndrome; EHR, electronic health records; TTE, transthoracic echocardiography.



Supplementary Figure 2: Kaplan-Meier mortality curves for patients with chronic coronary syndrome stratified by severity of valvular heart disease (VHD). There is a considerable overlap between the mortality curves of patients with moderate and severe valvular heart disease. Shaded areas represent the 95% confidence intervals.



Supplementary Figure 3: Kaplan-Meier mortality curves for patients with chronic coronary syndrome stratified by left ventricular dysfunction (LVD) and valvular heart disease (VHD). In this figure, VHD is moderate and severe combined VHD. Patients with VHD and LVD have higher mortality rates compared to patients with LVD only (Log-Rank p=0.01). Shaded areas represent the 95% confidence intervals.



Supplementary Figure 4: Kaplan-Meier mortality curves for patients with chronic coronary syndrome stratified by the number of valves affected. In this figure, valvular heart disease (VHD) is moderate and severe combined VHD. Shaded areas represent the 95% confidence intervals.

Supplementary Table 1: Data availability

Characteristics	All patients (n=1984)
Follow-up (years), n (%)	1984 (100%)
Age (years), n (%)	1984 (100%)
Male, n (%)	1984 (100%)
Hypertension, n (%)	1984 (100%)
Diabetes, n (%)	1984 (100%)
Dyslipidaemia, n (%)	1984 (100%)
Current or former smoker, n (%)	1984 (100%)
Family history of CAD, n (%)	1984 (100%)
Myocardial infarction, n (%)	1984 (100%)
PCI, n (%)	1984 (100%)
CABG, n (%)	1984 (100%)
SAVR, n (%)	1984 (100%)
TAVR, n (%)	1984 (100%)
SMVR, n (%)	1984 (100%)
TMVR, n (%)	1984 (100%)
STVR, n (%)	1984 (100%)
SPVR, n (%)	1984 (100%)
Atrial fibrillation/-flutter, n (%)	1984 (100%)
Stroke, n (%)	1984 (100%)
COPD, n (%)	1984 (100%)
Chest pain, n (%)	1984 (100%)
Dyspnea, n (%)	1984 (100%)
Other cardiac symptoms, n (%)	1984 (100%)
BMI (kg/m ²), mean (SD)	1634 (83%)
eGFR (ml/min/1.73 m ²), mean (SD)	1868 (94%)
Cholesterol (mmol/l), mean (SD)	1497 (75%)
LDL (mmol/l), mean (SD)	1231 (62%)
Triglyceride (mmol/l), mean (SD)	1454 (73%)
Antiplatelet therapy, n (%)	1984 (100%)
Anticoagulants, n (%)	1984 (100%)
ACE-inhibitor/ARB, n (%)	1984 (100%)
Beta-blockers, n (%)	1984 (100%)
Nitrates or other antianginal drugs, n (%)	1984 (100%)
Calcium antagonists, n (%)	1984 (100%)
Diuretics, n (%)	1984 (100%)
Statins, n (%)	1984 (100%)
Insulin, n (%)	1984 (100%)
Other oral diabetic drugs, n (%)	1984 (100%)

Availability of data. The total proportion of missing data was 3%.

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; CAD, coronary artery disease; eGFR, estimated glomerular filtration rate; LDL, low-density lipoprotein cholesterol; SAVR, surgical aortic valve replacement; SMVR, surgical mitral valve replacement; SPVR, surgical pulmonary valve replacement; STVR, surgical tricuspid valve replacement; TAVR, transcatheter aortic valve replacement; TEER, transcatheter edge-to-edge repair; VHD, valvular heart disease.

Supplementary Table 2: Baseline Characteristics of study population

Characteristics	All patients (n=1984)	No/mild VHD (n=1615)	Moderate VHD (n=325)	Severe VHD (n=44)	P-values
Follow-up (years), median [Q1,Q3]	3.49 [1.72, 5.62]	3.49 [1.72, 5.64]	3.46 [1.77, 5.60]	2.87 [1.29, 5.25]	0.501
Age (years), median [Q1,Q3]	65.00 [57.00, 73.00]	64.00 [56.00, 71.00]	72.00 [65.00, 80.00]	76.50 [69.75, 82.50]	<0.001
Male, n (%)	1169 (58.9)	959 (59.4)	187 (57.5)	23 (52.3)	0.548
Risk factors					
Hypertension, n (%)	1072 (54.0)	845 (52.3)	202 (62.2)	25 (56.8)	0.005
Diabetes, n (%)	516 (26.0)	429 (26.6)	77 (23.7)	10 (22.7)	0.494
Dyslipidaemia, n (%)	651 (32.8)	526 (32.6)	109 (33.5)	16 (36.4)	0.830
Current or former smoker, n (%)	695 (35.0)	585 (36.2)	105 (32.3)	5 (11.4)	0.002
Family history of CAD, n (%)	635 (32.0)	559 (34.6)	74 (22.8)	2 (4.5)	<0.001
Medical History					
Myocardial infarction, n (%)	510 (25.7)	397 (24.6)	104 (32.0)	9 (20.5)	0.015
PCI, n (%)	644 (32.5)	516 (32.0)	113 (34.8)	15 (34.1)	0.596
CABG, n (%)	207 (10.4)	145 (9.0)	60 (18.5)	2 (4.5)	<0.001
SAVR, n (%)	55 (2.8)	36 (2.2)	16 (4.9)	3 (6.8)	0.007

Continued on next page

Characteristics	All patients (n=1984)	No/mild VHD (n=1615)	Moderate VHD (n=325)	Severe VHD (n=44)	P-values
TAVR, n (%)	9 (0.5)	5 (0.3)	3 (0.9)	1 (2.3)	0.062
SMVR, n (%)	17 (0.9)	11 (0.7)	5 (1.5)	1 (2.3)	0.182
TEER, n (%)	1 (0.1)	0 (0.0)	0 (0.0)	1 (2.3)	<0.001
STVR, n (%)	6 (0.3)	6 (0.4)	0 (0.0)	0 (0.0)	0.503
SPVR, n (%)	1 (0.1)	1 (0.1)	0 (0.0)	0 (0.0)	0.892
Atrial fibrillation/-flutter, n (%)	246 (12.4)	131 (8.1)	93 (28.6)	22 (50.0)	<0.001
Stroke, n (%)	111 (5.6)	82 (5.1)	27 (8.3)	2 (4.5)	0.066
COPD, n (%)	128 (6.5)	90 (5.6)	31 (9.5)	7 (15.9)	0.001
Clinical examination					
Chest pain, n (%)	1168 (58.9)	986 (61.1)	164 (50.5)	18 (40.9)	<0.001
Dyspnea, n (%)	606 (30.5)	471 (29.2)	116 (35.7)	19 (43.2)	0.012
Other cardiac symptoms, n (%)	418 (21.1)	340 (21.1)	72 (22.2)	6 (13.6)	0.429
BMI (kg/m ²), mean (SD)	27.45 (5.39)	27.66 (5.40)	26.64 (5.28)	26.26 (5.19)	0.007
Laboratory parameters					
eGFR (ml/min/1.73 m ²) ^a , mean (SD)	70.05 (20.92)	72.02 (20.53)	61.71 (20.22)	62.07 (23.31)	<0.001
Cholesterol (mmol/l), mean (SD)	4.56 (1.27)	4.58 (1.26)	4.52 (1.36)	4.30 (1.27)	0.495
LDL (mmol/l), mean (SD)	2.56 (1.13)	2.58 (1.12)	2.48 (1.15)	2.45 (0.97)	0.450
Triglyceride (mmol/l), mean (SD)	1.62 (1.09)	1.64 (1.09)	1.51 (1.10)	1.37 (0.91)	0.113
Baseline medication					
Antiplatelet therapy, n (%)	1212 (61.1)	998 (61.8)	191 (58.8)	23 (52.3)	0.285
Anticoagulants, n (%)	294 (14.8)	176 (10.9)	98 (30.2)	20 (45.5)	<0.001
ACE-inhibitor/ARB, n (%)	901 (45.4)	705 (43.7)	172 (52.9)	24 (54.5)	0.004
Beta-blockers, n (%)	1035 (52.2)	797 (49.3)	214 (65.8)	24 (54.5)	<0.001
Nitrates or other antianginal drugs, n (%)	555 (28.0)	444 (27.5)	102 (31.4)	9 (20.5)	0.192
Calcium antagonists, n (%)	584 (29.4)	481 (29.8)	97 (29.8)	6 (13.6)	0.067
Diuretics, n (%)	501 (25.3)	364 (22.5)	119 (36.6)	18 (40.9)	<0.001
Statins, n (%)	1192 (60.1)	946 (58.6)	222 (68.3)	24 (54.5)	0.004
Insulin, n (%)	226 (11.4)	177 (11.0)	43 (13.2)	6 (13.6)	0.448
Other oral diabetic drugs, n (%)	365 (18.4)	308 (19.1)	50 (15.4)	7 (15.9)	0.268

Values are mean (standard deviation), n (%), or median [interquartile range].

^a Calculated with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation.

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; CAD, coronary artery disease; eGFR, estimated glomerular filtration rate; LDL, low-density lipoprotein cholesterol; SAVR, surgical aortic valve replacement; SMVR, surgical mitral valve replacement; SPVR, surgical pulmonary valve replacement; STVR, surgical tricuspid valve replacement; TAVR, transcatheter aortic valve replacement; TEER, transcatheter edge-to-edge repair; VHD, valvular heart disease.

Supplementary Table 3: Echocardiography characteristics of study population

Characteristics	All patients (n=1984)	No/mild VHD (n=1615)	Moderate VHD (n=325)	Severe VHD (n=44)
Left ventricular function, n (%)				
Normal	1479 (74.5)	1258 (77.9)	197 (60.6)	24 (54.5)
Mildly impaired	310 (15.6)	230 (14.2)	69 (21.2)	11 (25.0)
Moderately impaired	156 (7.9)	105 (6.5)	44 (13.5)	7 (15.9)
Severely impaired	39 (2.0)	22 (1.4)	15 (4.6)	2 (4.5)
Aortic stenosis grade, n (%)				
Normal/mild	1896 (95.6)	1615 (100.0)	262 (80.6)	19 (43.2)
Moderate	64 (3.2)	0 (0.0)	63 (19.4)	1 (2.3)
Severe	24 (1.2)	0 (0.0)	0 (0.0)	24 (54.5)
Aortic regurgitation grade, n (%)				
Normal/mild	1930 (97.3)	1615 (100.0)	279 (85.8)	36 (81.8)
Moderate	50 (2.5)	0 (0.0)	46 (14.2)	4 (9.1)
Severe	4 (0.2)	0 (0.0)	0 (0.0)	4 (9.1)
Mitral stenosis grade, n (%)				
Normal/mild	1978 (99.7)	1615 (100.0)	320 (98.5)	43 (97.7)
Moderate	6 (0.3)	0 (0.0)	5 (1.5)	1 (2.3)
Severe	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Mitral regurgitation grade, n (%)				
Normal/mild	1805 (91.0)	1615 (100.0)	164 (50.5)	26 (59.1)
Moderate	176 (8.9)	0 (0.0)	161 (49.5)	15 (34.1)
Severe	3 (0.2)	0 (0.0)	0 (0.0)	3 (6.8)
Tricuspid regurgitation grade, n (%)				
Normal/mild	1840 (92.7)	1615 (100.0)	204 (62.8)	21 (47.7)

Continued on next page

Characteristics	All patients (n=1984)	No/mild VHD (n=1615)	Moderate VHD (n=325)	Severe VHD (n=44)
Moderate	128 (6.5)	0 (0.0)	121 (37.2)	7 (15.9)
Severe	16 (0.8)	0 (0.0)	0 (0.0)	16 (36.4)
Pulmonary stenosis grade, n (%)				
Normal/mild	1983 (99.9)	1615 (100.0)	325 (100.0)	43 (97.7)
Moderate	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Severe	1 (0.1)	0 (0.0)	0 (0.0)	1 (2.3)
Pulmonary regurgitation grade, n (%)				
Normal/mild	1977 (96.6)	1615 (100.0)	320 (98.5)	42 (95.5)
Moderate	7 (0.4)	0 (0.0)	5 (1.5)	2 (4.5)
Severe	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

VHD = valvular heart disease.

Supplementary Table 4: Univariable and multivariable Cox regression analysis of variables associated with mortality: moderate and severe analysis

Variable	Univariable analysis		Multivariable analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
No/mild VHD	1 [Reference]	NA	1 [Reference]	NA
Moderate VHD	2.2 (1.7-2.8)	<0.001	1.4 (1.05-1.78)	0.02
Severe VHD	2.8 (1.6-4.8)	<0.001	0.99 (0.52-1.87)	0.96

Model was adjusted for age, diabetes, current or former smoking, valve repair or replacement, chronic obstructive pulmonary disease, chest pain, impaired renal function, and LV dysfunction. CI, confidence interval; HR, hazard ratio; VHD, valvular heart disease.

Supplementary Table 5: Univariable and multivariable Cox regression analysis of variables associated with mortality: valvular heart disease subtype analysis

Variable	Univariable analysis		Multivariable analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
AS				
No/mild AS	1 [Reference]	NA	1 [Reference]	NA
Moderate AS	1.5 (0.9-2.6)	0.09		
Severe AS	2.5 (1.3-5.2)	0.009		
AR				
No/mild AR	1 [Reference]	NA	1 [Reference]	NA
Moderate AR	1.8 (1.04-3.02)	0.04		
Severe AR	3.4 (0.9-13.8)	0.08		
MR				
No/mild MR	1 [Reference]	NA	1 [Reference]	NA
Moderate MR	2.3 (1.7-3.0)	<0.001		
Severe MR	4.2 (0.6-30.3)	0.15		
TR				
No/mild TR	1 [Reference]	NA	1 [Reference]	NA
Moderate TR	2.6 (1.9-3.6)	<0.001	1.6 (1.2-2.3)	0.005
Severe TR	2.5 (0.9-6.7)	0.07	1.2 (0.3-2.4)	0.8

Model was adjusted for age, diabetes, current or former smoking, valve repair or replacement, chronic obstructive pulmonary disease, chest pain, impaired renal function, and LV dysfunction. CI, confidence interval; HR, hazard ratio; VHD, valvular heart disease.

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

Mitchel A. Molenaar

Berto J. Bouma

Folkert W.A. Asselbergs

Niels J. Verouden

Jasper L. Selder

Steven A.J. Chamuleau

Mark J. Schuuring

Abstract

Background

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

Methods and results

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

Conclusion

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

Introduction

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

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

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

Methods

Training cohort

The training cohort was used to train the ML model. This cohort consisted of patients aged 18 years or older diagnosed with CCS who had both an outpatient visit and TTE at the Amsterdam University Medical Center (AUMC), location AMC, the Netherlands, between 2014 and 2017. Chronic coronary syndrome was defined as a clinical presentation of coronary artery disease at the outpatient visit, with the exception of patients in which acute coronary syndrome was the primary clinical presentation⁽¹⁾. The diagnosis was determined by the treating physician based on history taking, and patients were treated with medication accordingly, in accordance with the ESC guidelines⁽¹⁾. Additional testing was performed at the discretion of the physician, commonly in patients where medical treatment was ineffective or when there was a need to confirm or refute the CCS diagnosis. Patients were consecutively selected from electronic health records, and

their data were extracted from pseudonymized electronic health records and echocardiography reports, which was further described in the study of Molenaar et al⁽⁴⁾.

Testing cohort

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

Outcome

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

Echocardiography data

Chronic coronary syndrome patients underwent two-dimensional TTE with tissue Doppler imaging using various machines, including Vivid 9 (GE Vingmed Ultrasound AS, Horten, Norway) in the training cohort and Philips Epiq, Philips Affiniti, and Philips IE33 (Philips Medical Systems, Best, The Netherlands) in the testing cohort. These TTE assessments were performed by clinical technicians, who followed the recommendations of the ESC guidelines^(10, 11), European Association of Cardiovascular Imaging⁽¹²⁾, and standard operating procedure⁽¹³⁾. The TTE images were analysed with vendor-specific software including GE EchoPAC (GE Vingmed Ultrasound AS) in the training cohort and Xcelera (Philips Medical Systems) and TomTec 2D Cardiac Performance Analysis (Munich, Germany) in the testing cohort.

The initial TTE assessment was performed qualitatively by a clinical technician or cardiology resident in routine clinical practice, not performed specifically for this study. Semi-quantitative and quantitative measurements of atrial and ventricular dimensions, right ventricular (RV) function, and valve stenosis or regurgitation were obtained if indicated, especially if clinical decisions were based on these findings^(10, 12, 13). In apical two-chamber and fourchamber images, tracings of the LV endocardial borders were performed. The Simpson's biplane method was used to estimate the end-diastolic and end-systolic LV volumes, through which LV ejection fraction (LVEF) was calculated. Atrial and ventricular enlargements were defined as an increase in the size of their respective chambers. Specifically, left atrial enlargement was defined as a left atrial volume index $>34 \text{ mL/m}^2$. Right atrial enlargement was defined as a right atrial volume index $\geq 30 \text{ mL/m}^2$ for male patients and $\geq 28 \text{ mL/m}^2$ for female patients. Left ventricular enlargement was defined as a LV end-diastolic dimension $>58.4 \text{ mm}$ for male patients and $>52.2 \text{ mm}$ for female patients. Right ventricular enlargement was defined as a RV basal diameter $\geq 42 \text{ mm}$ ^(12, 14). The results were documented in a TTE report⁽¹²⁾, which was supervised by a dedicated imaging car-

diologist who examined the TTE images and made corrections to the TTE report if needed.

The TTE reports were de-identified, and one report was selected for each patient that was closest to the date of the outpatient visit. The following data were extracted from the TTE reports and recorded in the registry: left and right atrial and ventricular enlargements, LV and RV functions, and severity of aortic stenosis, aortic regurgitation, mitral stenosis, mitral regurgitation, tricuspid regurgitation, pulmonary stenosis, and pulmonary regurgitation. Left ventricular dysfunction was defined as mildly to severely impaired LV function ($LVEF \leq 51\%$ for male and $\leq 53\%$ for female). Moderately and severely impaired LV function was defined as an LVEF of 30–41% and <30%, respectively⁽¹⁵⁾. The estimated glomerular filtration rate (eGFR) was calculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation⁽¹⁶⁾.

Available data and standard risk scores

The following data were available for the training and testing cohorts and were used for model training: demographic data, cardiovascular risk factors, medical history, clinical examination, laboratory measurements, and echocardiographic data. A total of 43 features were used for training, of which 14 features were echocardiographic features. The Framingham risk score⁽¹⁷⁾ and ESC Systematic Coronary Risk Evaluation [SCORE2 (<70 years)/ SCORE2-OP (≥ 70 years)] risk^(18, 19) were calculated to estimate the 10-year cardiovascular risk for each patient. The variables included in these risk scores are age, gender (only in Framingham risk score), geographical region (only in SCORE2/SCORE2-OP), systolic blood pressure, diabetes (only in SCORE2/SCORE2/OP), smoking status, blood pressure treatment (only in Framingham risk score), and total-cholesterol and HDL-cholesterol.

Feature imputation and selection

Missing values in the training and testing cohorts were imputed by multiple imputation by chained equation with a linear regression model in 10 iterations. This process was repeated for the data of both the training and the testing cohorts separately. An eXtreme Gradient Boosting (XGBoost) model was trained to predict 5-year mortality. The eXtreme Gradient Boosting is a non-linear model that employs an ensemble of decision tree models, which has shown good performance in diverse classification problems^(20, 21). A grid search with five times five-fold cross-validation was conducted on the data of the training cohort to tune the hyperparameters of XGBoost (Supplementary Table 1). To obtain the optimal set of features and minimize the risk of overfitting, the tuned XGBoost model was trained using a 10 times 10-fold cross-validation strategy. This involved multiple training cycles in which the model was trained on 90% of the training data and validated on the remaining 10%. This process was repeated 10 times with other randomization of the data to obtain a reliable performance estimate. The feature that most frequently exhibited the lowest importance in the 10 iterations of cross-validation, as observed by the features' importance function of XGBoost, was excluded from the subsequent rounds of model training. To reduce the risk of overfitting and enhance the model's generalizability, the

minimal set of features was selected at the point where the model's performance began to decline.

Model training and testing

The ML model was trained with the selected features using 10-fold crossvalidation, with the area under the receiver operating characteristic curve (AUC) as the optimization metric. The model was recalibrated using the training data to further improve the calibration (agreement between the observed and the predicted risk of mortality) and was subsequently tested on the testing set. In the testing set, the model was evaluated a thousand times on randomly selected bootstrap samples to obtain a reliable estimate of performance. The ML model was trained using Python version 3.8 on a Windows-based computer (CPU: 2.3 GHz, RAM: 8 GB).

The impact of a feature on the ML model's outcome was further illustrated with Shapley Additive exPlanations (SHAP) plots⁽²²⁾. These plots provide insight into the decision-making process of the model, even for a specific patient, by illustrating both the importance and the impact direction (positive/negative) of a feature compared with a baseline value.

Statistical analysis

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

The ML-based risk score was compared with the performance of the following benchmark scores: (i) LV dysfunction; (ii) Framingham risk score⁽¹⁷⁾; (iii) SCORE2/SCORE2-OP^(18, 19); and (iv) a Cox-based risk score, a commonly used model in survival analysis⁽²³⁾. The AUC values of these scores were reported with corresponding 95% confidence intervals (CIs). The AUC of the ML model was compared with the AUC of the traditional risk scores according to DeLong's test⁽²⁴⁾. To derive the Cox-based risk score, a multivariable Cox survival analysis was performed on data in the training set using stepwise backward selection minimizing the Akaike information criterion. With this approach, the most significant predictors were selected in the final Cox model, and the inclusion of non-significant factors was avoided. The natural logarithm of the adjusted hazard ratio for each selected variable served as a coefficient in the Cox-based risk score. To calculate the Cox-based risk for each patient, the coefficients were multiplied by the corresponding data values, and the resulting products were summed. Precision-recall curves were plotted, which provide an insight into the performance of the model, especially when the number of patients in the classes (mortality/ non-mortality) are imbalanced.

The calibration of the models was assessed by plotting the predicted risk against the true proportion of mortality across multiple risk categories. In a calibrated model, the predicted proportion of mortality matches the true proportion in each category. The Brier score was calculated

as a quantitative measure of the accuracy of the model's predictions. The Brier score ranges between 0 (accurate prediction) and 1 (inaccurate prediction). The prognostic value of the ML score was further evaluated using survival analysis. A cut-off value was determined for the ML score at which the true positive rate minus the false-positive rate was maximal. Patients in the testing set with an ML score greater than the cut-off value were classified as high-risk patients, while patients with an ML score equal to or lower than the cut-off value were classified as low-risk patients. Kaplan–Meier curves were plotted to analyse the survival of patients categorized into high-risk and low-risk groups. The prognostic value of the ML score was compared with LV dysfunction by calculating the unadjusted hazard ratios and the logrank test.

This study meets all CODE-EHR minimum framework standards for the use of healthcare data for clinical research⁽²⁵⁾. Statistical analyses were performed in Python version 3.8 and RStudio version 2022.07.0 (RStudio Team, Boston, MA, USA) using R-version 4.1.3 (R Core Team, Vienna, Austria). A P-value <0.05 was considered statistically significant.

Results

Study population

A total of 2845 patients with CCS were screened in both tertiary centres. Among them, 1988 patients had an outpatient visit and underwent TTE between 2014 and 2021. Patients who were excluded due to the absence of a TTE assessment ($n = 535$) had baseline characteristics comparable with patients with a TTE assessment (see Supplementary Table 2). The only significant difference was in age, which was higher in the group without a TTE assessment (68 vs. 65 years, $P < 0.001$). After an exclusion of 735 patients who did not have an outpatient visit before 2017 ($n = 731$) or did not have complete TTE acquisition ($n = 4$), 1253 patients were included (775 patients in the training set and 478 patients in the testing set). The flowchart is depicted in Figure 1.

Baseline characteristics and follow-up

The baseline characteristics of the patients in the training and testing sets are shown in Tables 1 and 2, respectively. Patients in the training set had a median age of 66 years, and 58% of them were male. Additional invasive coronary angiography was most frequently (60%) performed in the diagnostic process, followed by a myocardial perfusion scan (39%) and computed tomography (CT) coronary angiography (23%). Hypertension was the most common risk factor (60%), followed by smoking (36%) and dyslipidaemia (34%). Prior myocardial infarction was reported in 206 patients (27%). Acute coronary syndrome was reported in 11 patients (1%) during the 3 months preceding the presentation at the outpatient clinic. Revascularization was performed by percutaneous coronary intervention in 244 patients (32%) and by coronary artery bypass grafting (CABG) in 75 patients (10%). Patients had a mean body mass index (BMI) of 28 kg/m^2 , and chest pain was reported in 427 patients (55%). Left ventricular dysfunction was reported in 161 patients (21%), with severe dysfunction in 6 (1%) patients. During the 5-year follow-up period,

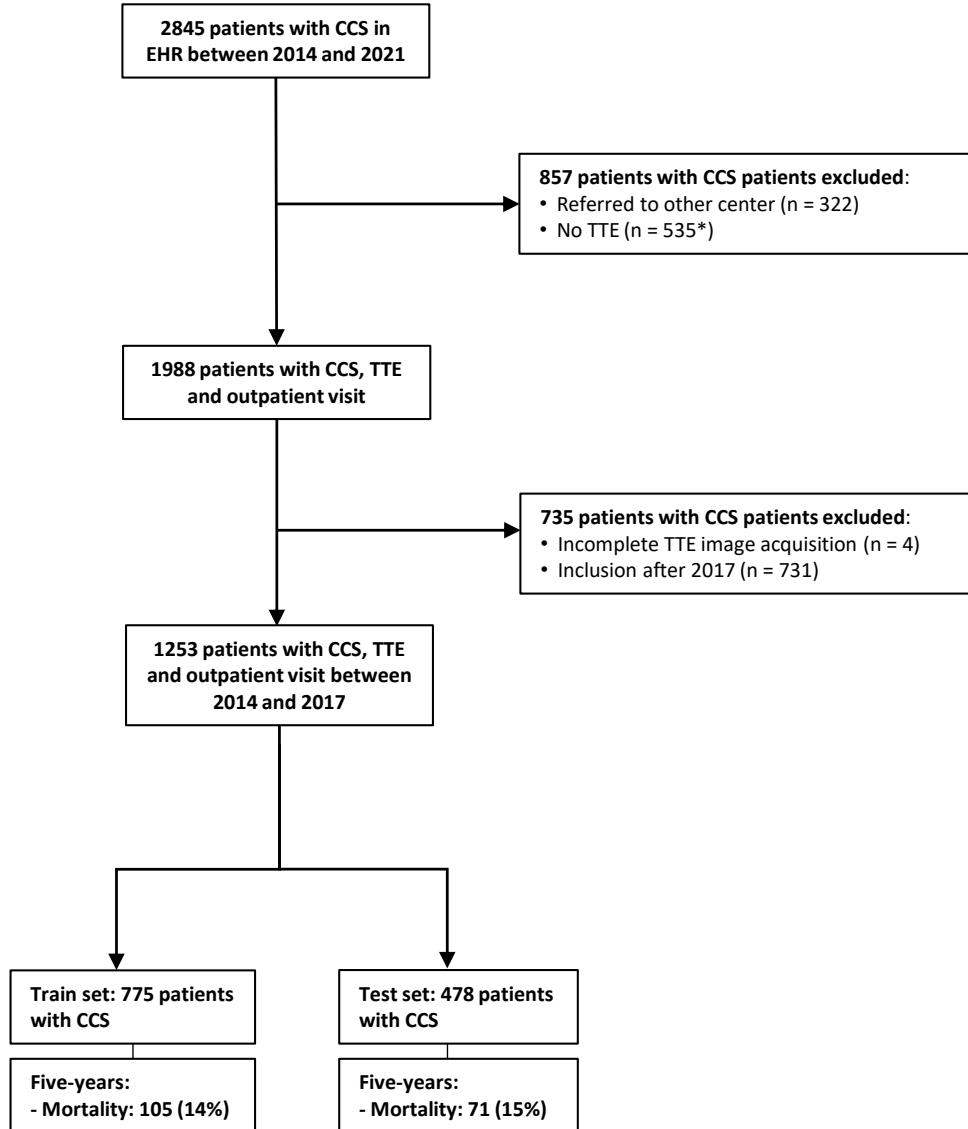


Figure 1: Inclusion flowchart. EHR, electronic health records; CCS, chronic coronary syndrome; TTE, transthoracic echocardiography. *The clinical characteristics of patients without transthoracic echocardiography (n = 535) and patients with complete transthoracic echocardiography (n = 1984) were comparable, as shown in Supplementary Table 2.

199 patients received revascularization (26%), 18 patients received valvular heart repair or replacement (2%), and 105 (14%) patients died (Table 3). A total of 2% of the data were missing, as shown in Supplementary Table 3.

Table 1: Baseline characteristics of patients in the training set

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

Continued on next page

Characteristics	All patients (n=775)	Alive (n=670)	Dead (n=105)	p-value
Moderate or severe aortic regurgitation, n (%)	19 (2.5)	14 (2.1)	5 (4.8)	0.191
Moderate or severe mitral regurgitation, n (%)	80 (10.3)	60 (9.0)	20 (19.0)	0.003
Moderate or severe mitral stenosis, n (%)	2 (0.3)	2 (0.3)	0 (0.0)	1.000
Moderate or severe pulmonary regurgitation, n (%)	4 (0.5)	2 (0.3)	2 (1.9)	0.161
Moderate or severe pulmonary stenosis, n (%)	1 (0.1)	1 (0.1)	0 (0.0)	1.000
Moderate or severe tricuspid regurgitation, n (%)	56 (7.2)	37 (5.5)	19 (18.1)	<0.001
Diagnostic modalities				
Exercise ECG, n (%)	133 (17.2)	123 (18.4)	10 (9.5)	0.036
Stress CMR, n (%)	61 (7.9)	54 (8.1)	7 (6.7)	0.766
Myocardial perfusion scan, n (%)	298 (38.5)	259 (38.7)	39 (37.1)	0.850
CT coronary angiography, n (%)	176 (22.7)	147 (21.9)	29 (27.6)	0.244
Invasive coronary angiography, n (%)	463 (59.7)	394 (58.8)	69 (65.7)	0.217
Baseline medication				
Antiplatelet therapy, n (%)	516 (66.6)	442 (66.0)	74 (70.5)	0.424
Anticoagulants, n (%)	102 (13.2)	77 (11.5)	25 (23.8)	0.001
ACE-inhibitor/ARB, n (%)	387 (49.9)	327 (48.8)	60 (57.1)	0.138
Beta-blockers, n (%)	468 (60.4)	398 (59.4)	70 (66.7)	0.191
Nitrates or other antianginal drugs, n (%)	211 (27.2)	178 (26.6)	33 (31.4)	0.356
Calcium antagonists, n (%)	249 (32.1)	213 (31.8)	36 (34.3)	0.692
Diuretics, n (%)	219 (28.3)	172 (25.7)	47 (44.8)	<0.001
Statins, n (%)	490 (63.2)	422 (63.0)	68 (64.8)	0.809
Insulin, n (%)	95 (12.3)	74 (11.0)	21 (20.0)	0.015
Other oral diabetic drugs, n (%)	179 (23.1)	151 (22.5)	28 (26.7)	0.419

ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blocker; BMI, body mass index; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; CT, computed tomography; ECG, electrocardiography; eGFR, estimated glomerular filtration rate; PCI, percutaneous coronary intervention; VHD, valvular heart disease.

^aWithin 3 months preceding the presentation at the outpatient clinic.

^bCalculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation.

Table 2: Baseline characteristics of patients in the training set and testing set

Characteristics	Training set (n=775)	Testing set (n=478)	p-value
Follow-up (years), median [Q1,Q3]	6.46 [5.48, 6.89]	6.34 [5.35, 6.92]	0.403
Age (years), median [Q1,Q3]	66.00 [58.00, 73.00]	66.00 [58.00, 75.00]	0.208
Male, n (%)	445 (57.4)	290 (60.7)	0.282
Risk factors			
Hypertension, n (%)	462 (59.6)	246 (51.5)	0.006
Diabetes, n (%)	232 (29.9)	122 (25.5)	0.105
Dyslipidaemia, n (%)	260 (33.5)	127 (26.6)	0.011
Current or former smoker, n (%)	280 (36.1)	179 (37.4)	0.682
Family history of CAD, n (%)	257 (33.2)	155 (32.4)	0.836
Medical history			
Myocardial infarction, n (%)	206 (26.6)	134 (28.0)	0.620
Recent ACS event ^a , n (%)	11 (1.4)	6 (1.2)	0.404
PCI, n (%)	244 (31.5)	190 (39.7)	0.003
CABG, n (%)	75 (9.7)	68 (14.2)	0.018
Valvular repair or replacement, n (%)	27 (3.5)	14 (2.9)	0.709
Atrial fibrillation/-flutter, n (%)	88 (11.4)	53 (11.1)	0.958
Stroke, n (%)	40 (5.2)	25 (5.2)	1.000
COPD, n (%)	58 (7.5)	32 (6.7)	0.680
Peripheral arterial disease, n (%)	40 (5.2)	17 (3.6)	0.236
Clinical examination			
Chest pain, n (%)	427 (55.1)	293 (61.3)	0.036
Dyspnea, n (%)	203 (26.2)	169 (35.4)	0.001
Other cardiac symptoms, n (%)	156 (20.1)	101 (21.1)	0.723
Systolic blood pressure (mmHg), mean (SD)	141.02 (21.79)	137.66 (22.67)	0.029
Heart rate (bpm), mean (SD)	71.74 (13.81)	70.27 (13.92)	0.130
BMI (kg/m ²), mean (SD)	27.89 (5.64)	28.57 (5.51)	0.063
eGFR (ml/min/1.73 m ²) ^b , mean (SD)	69.34 (21.00)	67.46 (20.83)	0.130
Total cholesterol (mmol/l), mean (SD)	4.57 (1.22)	4.50 (1.22)	0.414
HDL (mmol/l), mean (SD)	1.32 (0.44)	1.35 (0.49)	0.317
LDL (mmol/l), mean (SD)	2.93 (1.14)	2.78 (1.11)	0.063
Triglyceride (mmol/l), mean (SD)	1.55 (1.13)	1.69 (0.96)	0.054
Echocardiographic measurement			
Left ventricular function, n (%)			<0.001
Normal	614 (79.2)	301 (63.0)	
Mildly impaired	109 (14.1)	105 (22.0)	
Moderately impaired	46 (5.9)	54 (11.3)	
Severely impaired	6 (0.8)	18 (3.8)	
Right ventricular dysfunction, n (%)	52 (6.7)	59 (12.3)	0.001
Right ventricular enlargement, n (%)	28 (3.6)	18 (3.8)	1.000
Left ventricular enlargement, n (%)	90 (11.6)	27 (5.6)	0.001
Left atrial enlargement, n (%)	338 (43.6)	111 (23.2)	<0.001
Right atrial enlargement, n (%)	91 (11.7)	28 (5.9)	0.001
Moderate or severe aortic stenosis, n (%)	33 (4.3)	27 (5.6)	0.325
Moderate or severe aortic regurgitation, n (%)	19 (2.5)	17 (3.6)	0.335
Moderate or severe mitral regurgitation, n (%)	80 (10.3)	31 (6.5)	0.026
Moderate or severe mitral stenosis, n (%)	2 (0.3)	3 (0.6)	0.585
Moderate or severe pulmonary regurgitation, n (%)	4 (0.5)	0 (0.0)	0.290
Moderate or severe pulmonary stenosis, n (%)	1 (0.1)	0 (0.0)	1.000
Moderate or severe tricuspid regurgitation, n (%)	56 (7.2)	31 (6.5)	0.699
Diagnostic modalities			

Continued on next page

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

ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blocker; BMI, body mass index; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; CT, computed tomography; ECG, electrocardiography; eGFR, estimated glomerular filtration rate; PCI, percutaneous coronary intervention; VHD, valvular heart disease.

^aWithin 3 months preceding the presentation at the outpatient clinic.

^bCalculated with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation.

Table 3: Baseline characteristics of patients in the training set and testing set

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

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

Feature selection

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

Prediction of mortality

The ML model [AUC: 0.79 (95% CI 0.78–0.81)] demonstrated superior discriminative performance compared with LV dysfunction [AUC: 0.64 (95% CI 0.63–0.66)], Framingham risk score [AUC: 0.62 (95% CI 0.60–0.63)], SCORE2/SCORE2-OP [AUC: 0.67 (95% CI 0.65–0.68)], and Cox-based risk score [AUC: 0.76 (95% CI 0.75–0.78), all P < 0.001]. The discriminative performance of

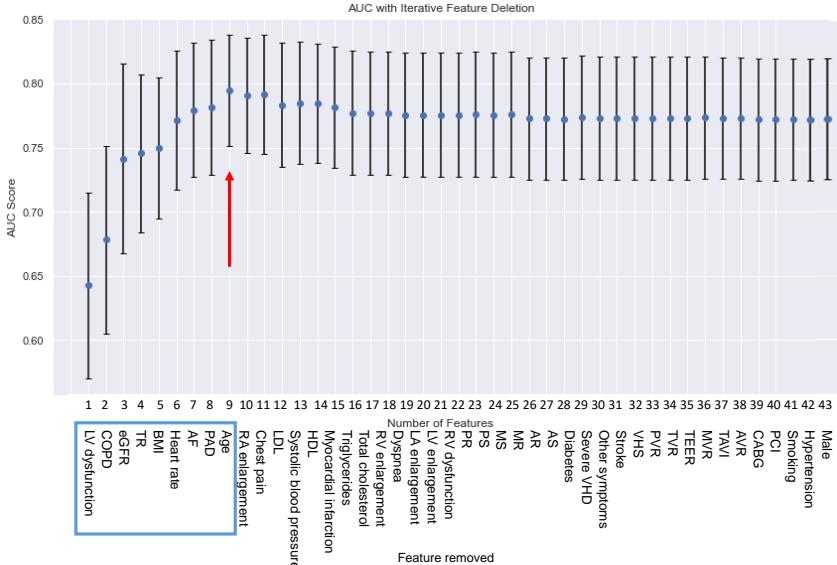


Figure 2: Number of features versus discriminative performance of machine learning model. The first model was trained with all features shown on the x-axis using a 10 times 10-fold cross-validation strategy on the training set. The feature that most frequently exhibited the lowest importance in the 10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

the models is shown in Figure 3. Compared with the traditional risk scores, the precision-recall curve of the ML model showed the best trade-off between precision and recall across different classification thresholds (see Supplementary Figure 1). The variables in the final Cox-based model included age, diabetes, current or former smoker, chronic obstructive pulmonary disease, chest pain, renal function ($eGFR < 60 \text{ mL/min/ } 1.73 \text{ m}^2$), LV dysfunction (mildly to severely impaired function), and atrial fibrillation/flutter.

External validation

The baseline characteristics and follow-up of patients in the testing set are shown in Tables 2 and 3, respectively. Patients in the testing set had comparable characteristics with those in the training set in terms of age, gender, and laboratory parameters. In the diagnostic process of patients in the testing set, additional invasive coronary angiography (71 vs. 60%) and CT coronary an-

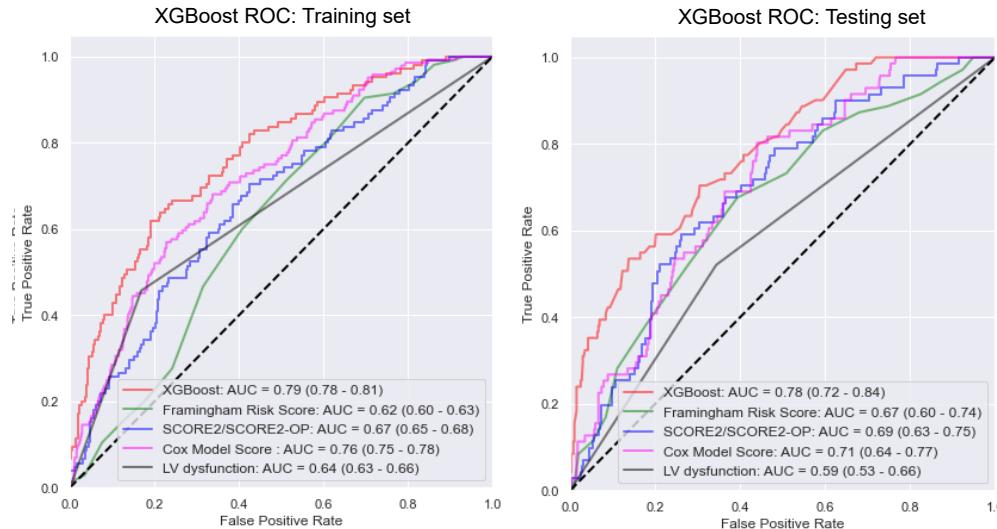


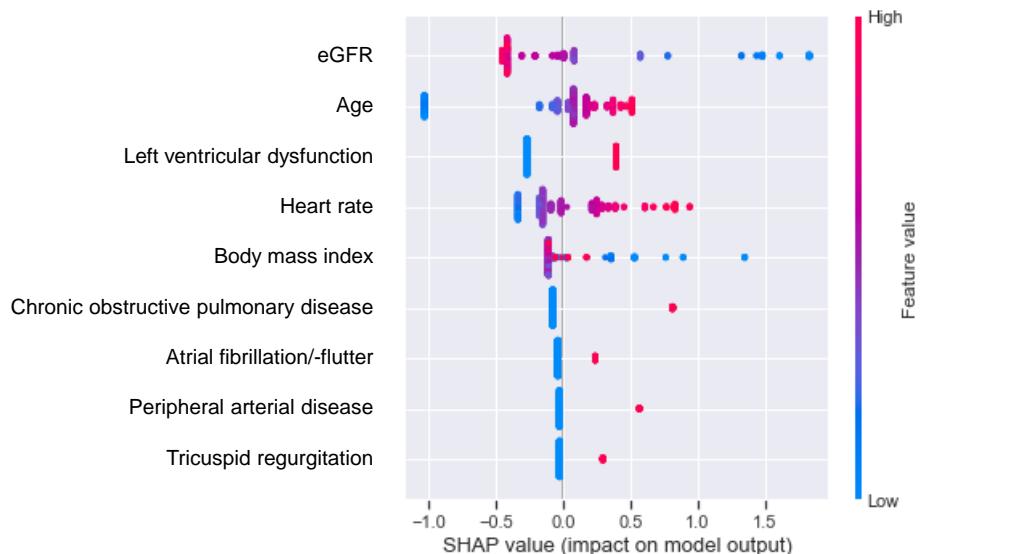
Figure 3: Discriminative performance of the machine learning model and traditional risk scores. The false-positive rate is plotted against the true positive rate across a range of classification thresholds for the machine learning model (eXtreme Gradient Boosting) and traditional risk scores (training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

giography (31 vs. 23%) were more frequently performed compared with the training set. Stress cardiac magnetic imaging was performed more frequently in the training set (2 vs. 8%). Patients in the testing set had lower rates of hypertension (52 vs. 60%) and dyslipidaemia (27 vs. 34%) compared with the training set, but higher rates of revascularization (percutaneous coronary intervention: 40 vs. 32%, CABG: 14 vs. 10%) and LV dysfunction (37 vs. 21%). Revascularization was more often reported in patients in the testing set during the 5-year follow-up period (36 vs. 26%). Mortality was reported in 71 patients (15%) in the testing set, which was comparable with patients in the training set. A total of 4% of the data were missing, as shown in Supplementary Table 3.

In external validation, the highest AUC was observed for the ML model [AUC: 0.78 (95% CI 0.72–0.84)], followed by the Cox-based score [AUC: 0.71 (95% CI 0.64–0.77), $P = 0.002$], SCORE2/SCORE2-OP [AUC: 0.69 (95% CI 0.63–0.75), $P = 0.03$], Framingham risk score [AUC: 0.67 (95% CI 0.60–0.74), $P = 0.006$], and LV dysfunction [AUC: 0.59 (95% CI 0.53–0.66), $P < 0.001$; Figure 3].

Individual risk prediction: explainable machine learning

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



3

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

Calibration

The predicted probabilities from the ML model were aligned with the observed mortality rates in each risk category. The Brier scores were 0.08 (95% CI 0.07–0.10) for the training set and 0.10 (95% CI 0.09–0.12) for the testing set, as shown in Supplementary Figure 2.

Risk stratification by the machine learning model

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

Discussion

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

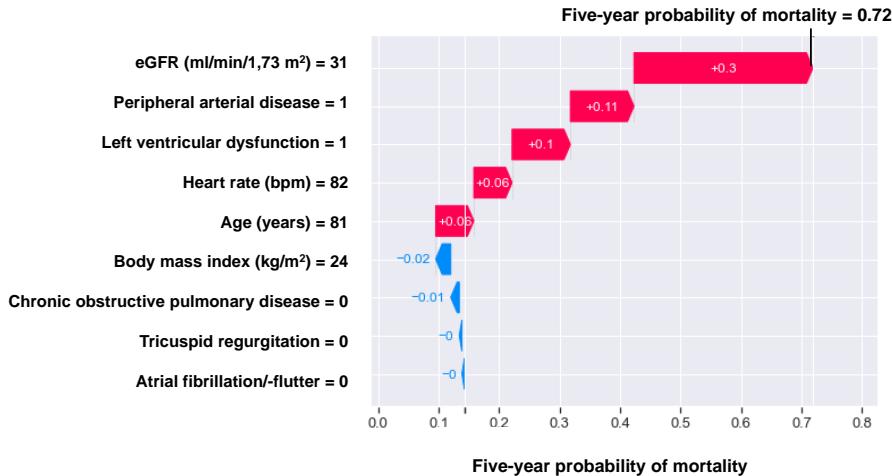
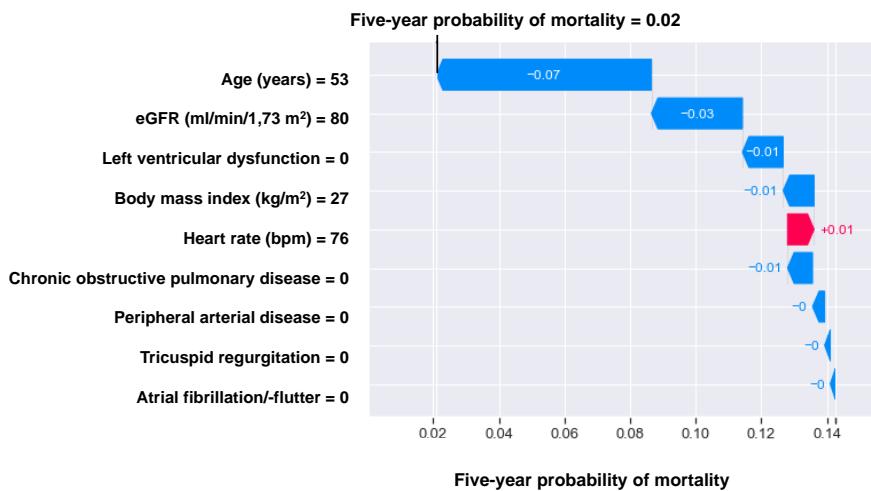
81-year-old male who died after a one-year follow-up period**54-year-old female who was alive after five-year follow-up period**

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

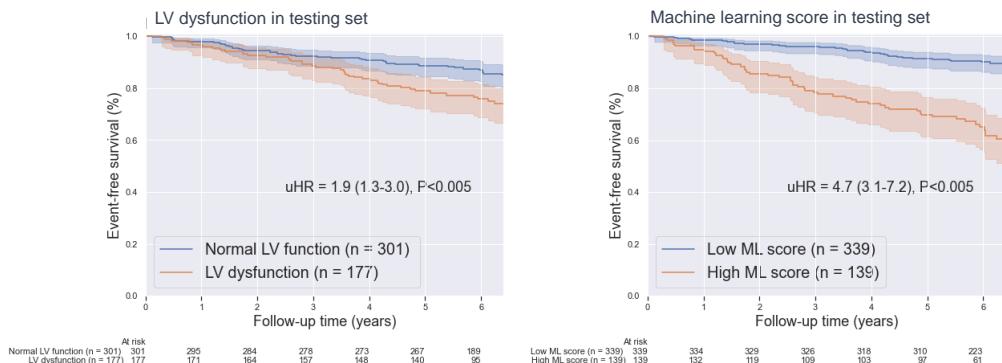


Figure 6: Kaplan–Meier survival curves for the machine learning score and LV dysfunction. The machine learning score (right figure) had a superior discriminative ability compared with left ventricular dysfunction (left figure) in distinguishing individuals at high and low risks of mortality. This score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

Machine learning for risk stratification

The prevalence and mortality rates of CCS are exhibiting an upward trend, posing a major challenge for risk stratification of these patients⁽²⁶⁾. To allocate healthcare resources to patients with a greater risk of cardiovascular events, accurate risk stratification is essential. However, in our study, risk stratification utilizing the gold standard LV function demonstrated a low predictive power. The incorporation of multiple variables as observed in the Framingham risk score, SCORE2/SCORE2-OP, and Cox model score resulted in more accurate discrimination compared with LV function, but not to the same extent as the ML model. In external validation, the ML model exhibited the highest discriminative performance, which suggests that risk stratification of CCS patients necessitates models that can incorporate non-linear relationships and complex interactions.

To our knowledge, this is the first study that investigated the performance of explainable ML in patients with CCS using TTE and clinical data. The use of TTE data in prediction models offers several advantages since it is widely used, non-invasive, and inexpensive⁽³⁾. Previous studies have included data from other imaging modalities in their ML models, which showed performances in line with our findings. Motwani et al.⁽⁸⁾ trained an ML model to predict 5-year mortality in 10030 patients with suspected CCS who underwent CCTA. Both clinical and CCTA data were included in the ML model, which exhibited a higher AUC (0.79) compared with the Framingham risk score and CCTA severity scores. More recently, Pezel et al.⁽⁷⁾ trained an ML model to predict mortality using clinical and stress magnetic resonance (CMR) in 31752 patients with suspected or known CCS. The authors showed that the ML model was able to predict 10-year mortality more accurately (AUC 0.76) compared with traditional risk scores. These studies emphasize the potential role of ML in addressing the challenge of overseeing the growing number of imaging and clinical

variables for risk assessment of the individual patient⁽⁷⁾.

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

Explainable machine learning

In this study, a tree-based ML model was trained that provides information about the model's decision-making by means of SHAP values. To date, there is an ongoing debate whether currently available explainability techniques are sufficient and a prerequisite to inform clinicians about decisions for the individual patient⁽²⁷⁾. Explanations regarding the model's output are only approximations, and the underlying model may be incorrect, which introduces bias in explainability techniques. Despite these acknowledged limitations, it is important to recognize that the explainability technique used in this study is one of the most effective approaches currently available to provide an insight into the decision-making process of the model⁽²⁸⁾. To further enhance trustworthiness, further efforts are needed to provide information about the certainty of a decision.

Study limitations

Several remarks can be made about this study. Patients without a TTE assessment at the investigated centres were excluded from analysis, which may have led to an unrepresentative CCS population. However, these patients had comparable baseline characteristics with those with TTE assessment, which reduces the risk of selection bias. The TTE assessments in this study were conducted as part of standard clinical practice. The reproducibility of these TTE assessments was not evaluated. In our study, BMI was chosen as an established risk factor as recommended by the ESC guidelines, despite the growing evidence that waist-to-height ratio may be a more accurate predictor of mortality^(1, 29–31).

The developed ML model was trained to predict all-cause 5-year mortality. The cause of mortality could not be obtained for all patients and was therefore not further differentiated. There is currently a lack of traditional 5-year mortality risk scores for patients with CCS. Therefore, we did the same as Motwani et al.⁽⁸⁾ did, who compared the ML score with traditional risk scores that are designed to estimate the 10-year risk of cardiovascular events. The number of patients included in this study was limited to patients from two tertiary centres and was low compared with that in previous ML studies. The strength of our study was the testing of the ML model in an external centre, which is a prerequisite to determine the generalizability of a prediction model⁽³²⁾. To prevent the model from overfitting, we iteratively eliminated the least important features until the model's performance began to decline. The final ML model generalized well (AUC testing set 0.78) on unseen data and outperformed the LV function and other traditional risk scores. Despite the lack of quantitative parameters of TTE (e.g. parameters of diastolic function)

and the limited number of variables in the ML model, its performance remained consistent in the testing set even when there were minor differences in patient characteristics compared with those in the training set. Datasets with a larger number of patients and the inclusion of quantitative parameters of TTE may facilitate the identification of more complex patterns related to mortality, which may further enhance the risk prediction in patients with CCS.

Our study primarily focused on the potential of ML-based mortality prediction using clinical and TTE data, and a comparison between this ML-TTE score and ML-based mortality prediction models on CCTA or CMR data was not within the scope of our research. A comparative analysis between ML-based TTE prediction and ML-based CCTA and CMR scores could provide insights into their relative performance for mortality prediction. In addition, combining these multimodal data in ML models may further enhance risk prediction. Once such a model is implemented in a clinical setting and prospectively validated; the effectiveness of ML-based risk stratification could be evaluated in randomized controlled trials⁽³³⁾.

Conclusions

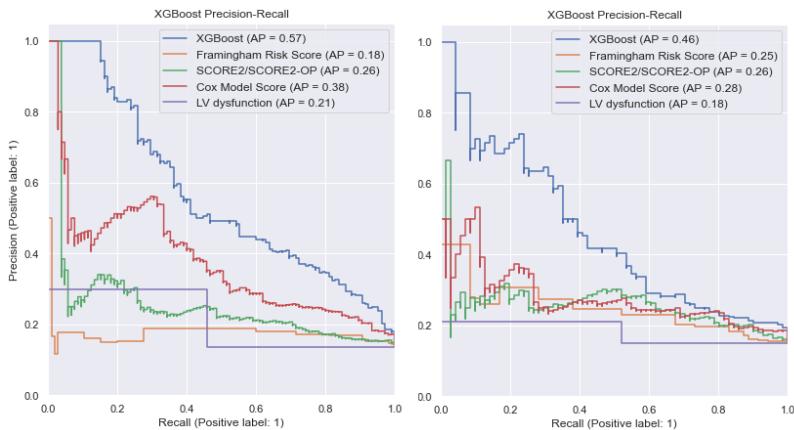
This study demonstrates that an explainable ML model using TTE and clinical data can accurately identify CCS patients with a high risk of 5-year mortality, with a prognostic value superior to those of LV dysfunction and other traditional risk scores. These findings are of clinical relevance since they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

References

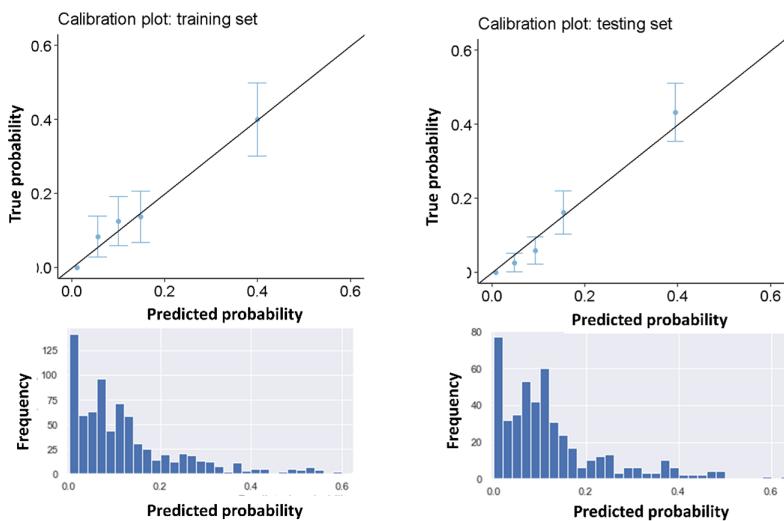
- [1] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;41:407–477.
- [2] Sorbets E, Fox KM, Elbez Y, Danchin N, Dorian P, Ferrari R, et al. Long-term outcomes of chronic coronary syndrome worldwide: insights from the international CLARIFY registry. *Eur Heart J* 2020;41:347–356.
- [3] Schuurings MJ, Işgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;8:e648877.
- [4] Molenaar MA, Bouma BJ, Coerkamp CF, Man JP, Işgum I, Verouden NJ, et al. The impact of valvular heart disease in patients with chronic coronary syndrome. *Front Cardiovasc Med* 2023;10:1211322.
- [5] Gill SK, Karwath A, Uh H-W, Cardoso VR, Gu Z, Barsky A, et al. Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *Eur Heart J* 2023; 44:713–725.
- [6] Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, et al. Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease. *Curr Cardiol Rep* 2022;24:365–376.
- [7] Pezel T, Sanguineti F, Garot P, Unterseeh T, Champagne S, Toupin S, et al. Machine-learning score using stress CMR for death prediction in patients with suspected or known CAD. *JACC Cardiovasc Imaging* 2022;15:1900–1913.
- [8] Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J* 2017; 38:500–507.
- [9] Mishra RK, Tison GH, Fang Q, Scherzer R, Whooley MA, Schiller NB. Association of machine learning-derived phenogroupings of echocardiographic variables with heart failure in stable coronary artery disease: the heart and soul study. *J Am Soc Echocardiogr* 2020;33:322–331.e1.
- [10] Vahanian A, Alfieri O, Andreotti F, Antunes MJ, Barón-Esquivias G, Baumgartner H, et al. Guidelines on the management of valvular heart disease (version 2012). *Eur Heart J* 2012;33:2451–2496.
- [11] Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, et al. 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J* 2017;38: 2739–2791.
- [12] Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2017;18:1301–1310.
- [13] Bouma BJ, Riezenbos R, Voogel AJ, Veldhorst MH, Jaarsma W, Hrudova J, et al. Appropriate use criteria for echocardiography in the Netherlands. *Neth Heart J* 2017; 25:330–334.
- [14] Koster J, Pieper PG. Normaalwaarden. In: Hamer JPM, Pieper PG, eds. Praktische echocardiografie. 3rd ed. Houten: Bohn Stafleu van Loghum; 2015, p389–410.
- [15] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 2015;28:1–39.e14.
- [16] Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. New equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–612.
- [17] D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–753.
- [18] SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;42:2439–2454.
- [19] SCORE2-OP Working Group and ESC Cardiovascular Risk Collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *Eur Heart J* 2021;42:2455–2467.
- [20] Al'Aref SJ, Singh G, Choi JW, Xu Z, Maliakal G, van Rosendaal AR, et al. A boosted ensemble algorithm for determination of plaque stability in high-risk patients on coronary CTA. *JACC Cardiovasc Imaging* 2020;13:2162–2173.
- [21] Wenzl FA, Krämer S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *Lancet* 2022;400:744–756.
- [22] Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 2017, p.4768–4777.
- [23] Cox DR. Regression models and life-tables. *J R Stat Soc B* 1972;34:187–202.
- [24] 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.
- [25] Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *Eur Heart J* 2022;43:3578–3588.
- [26] Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J Am Coll Cardiol* 2020;76:2982–3021.
- [27] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745–e750.

- [28] Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021;113:103655.
- [29] Ashwell M, Mayhew L, Richardson J, Rickayzen B. Waist-to-height ratio is more predictive of years of life lost than body mass index. *PLoS One* 2014;9:e103483.
- [30] Abdi Dezfooli R, Mohammadian Khonsari N, Hosseinpour A, Asadi S, Ejtahed H-S, Qorbani M. Waist to height ratio as a simple tool for predicting mortality: a systematic review and meta-analysis. *Int J Obes* 2023;47:1286–1301.
- [31] Cox BD, Whichelow M. Ratio of waist circumference to height is better predictor of death than body mass index. *BMJ* 1996;313:1487.
- [32] Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derkken-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826–832.
- [33] Schuring M, Man J, Chamuleau S. Inclusive health tracking: unlock the true potential of digital health solutions. *JACC: Advances* 2023;2:100545.

Supplementary material



Supplementary Figure 1: Precision-recall curves. The machine learning model (XGBoost) demonstrated the best trade-off between precision and recall across different classification thresholds compared with the traditional risks scores in both the training set (left Figure) and the testing set (right Figure). AP, average precision; LV, left ventricular; XGBoost, eXtreme Gradient Boosting.



Supplementary Figure 2: Calibration and frequency plots for machine learning model (XGBoost) are shown for the training set (left Figures) and testing set (right Figures). The calibration plots depict the actual frequency of mortality and the predicted proportion of mortality, grouped by a decile of risk. The model exhibited good calibration when evaluated on the training set (Brier score: 0.08 [95% CI 0.07-0.10]) and the testing set (Brier score: 0.10 [95% CI 0.09-0.12]).

Supplementary Table 1: Hyperparameter tuning of XGBoost model

Hyperparameter	Options	Included in final model
Learning rate (eta)	[0.01, 0.03, 0.05, 0.07, 0.1]	0.05
Number of decision trees (n_estimators)	[100, 200, 300, 400]	200
Maximum depth of decision tree (max_depth)	[1,2,3,4]	1
Fraction of training data used to train each individual tree (subsample)	[0.7,0.8,0.9]	0.7

Supplementary Table 2: Baseline characteristics of patients with and without transthoracic echocardiography

Characteristics	Patients with TTE (n=1984)	Patients without TTE (n=535)	p-value
Age (years), median [Q1,Q3]	65.0 [57.0, 73.0]	68.0 [61.0, 75.0]	<0.001
Male, n (%)	1169 (58.9)	319 (60.0)	0.54
Systolic blood pressure (mmHg), mean (SD)	139.3 (22.0)	137.2 (20.8)	0.103
Heart rate (bpm), mean (SD)	71.6 (14.2)	70.8 (14.8)	0.412
BMI (kg/m^2), mean (SD)	27.5 (5.4)	27.7 (4.7)	0.476
eGFR (ml/min/1.73 m^2), mean (SD)	70.0 (20.9)	68.8 (18.1)	0.218

BMI, body mass index; eGFR, estimated glomerular filtration rate; TTE, transthoracic echocardiography.

Supplementary Table 3: Data availability

Characteristics	Train set	Test set
Follow-up (years), n (%)	775 (100%)	478 (100%)
Age (years), n (%)	775 (100%)	478 (100%)
Male, n (%)	775 (100%)	478 (100%)
Risk factors		
Hypertension, n (%)	775 (100%)	478 (100%)
Diabetes, n (%)	775 (100%)	478 (100%)
Dyslipidaemia, n (%)	775 (100%)	478 (100%)
Current or former smoker, n (%)	775 (100%)	478 (100%)
Family history of CAD, n (%)	775 (100%)	478 (100%)
Medical History		
Myocardial infarction, n (%)	775 (100%)	478 (100%)
PCI, n (%)	775 (100%)	478 (100%)
CABG, n (%)	775 (100%)	478 (100%)
Valvular repair or replacement, n (%)	775 (100%)	478 (100%)
Atrial fibrillation/-flutter, n (%)	775 (100%)	478 (100%)
Stroke, n (%)	775 (100%)	478 (100%)
COPD, n (%)	775 (100%)	478 (100%)
Peripheral arterial disease, n (%)	775 (100%)	478 (100%)
Clinical examination		
Chest pain, n (%)	775 (100%)	478 (100%)
Dyspnea, n (%)	775 (100%)	478 (100%)
Other cardiac symptoms, n (%)	775 (100%)	478 (100%)
Systolic blood pressure (mmHg), mean (SD)	650 (84%)	307 (64%)
Heart rate (bpm), mean (SD)	642 (83%)	297 (62%)
BMI (kg/m^2), mean (SD)	630 (81%)	269 (56%)
Echocardiographic measurement		
Left ventricular function, n (%)	775 (100%)	478 (100%)
Right ventricular dysfunction, n (%)	775 (100%)	478 (100%)
Right ventricular enlargement, n (%)	775 (100%)	478 (100%)
Left ventricular enlargement, n (%)	775 (100%)	478 (100%)
Left atrial enlargement, n (%)	775 (100%)	478 (100%)
Right atrial enlargement, n (%)	775 (100%)	478 (100%)
Moderate or severe aortic stenosis, n (%)	775 (100%)	478 (100%)
Moderate or severe aortic regurgitation, n (%)	775 (100%)	478 (100%)
Moderate or severe mitral regurgitation, n (%)	775 (100%)	478 (100%)
Moderate or severe mitral stenosis, n (%)	775 (100%)	478 (100%)
Moderate or severe pulmonary regurgitation, n (%)	775 (100%)	478 (100%)
Moderate or severe pulmonary stenosis, n (%)	775 (100%)	478 (100%)
Moderate or severe tricuspid regurgitation, n (%)	775 (100%)	478 (100%)
Laboratory parameters		

Continued on next page

Characteristics	Train set	Test set
eGFR (ml/min/1.73 m ²), mean (SD)	753 (97%)	460 (96%)
Total cholesterol (mmol/l), mean (SD)	642 (83%)	345 (72%)
HDL (mmol/l), mean (SD)	624 (81%)	317 (66%)
LDL (mmol/l), mean (SD)	624 (81%)	317 (66%)
Triglyceride (mmol/l), mean (SD)	638 (82%)	326 (68%)
Diagnostic modalities		
Exercise ECG, n (%)	683 (88%)	415 (87%)
Stress CMR, n (%)	683 (88%)	415 (87%)
Myocardial perfusion scan, n (%)	683 (88%)	415 (87%)
CT coronary angiography, n (%)	683 (88%)	415 (87%)
Invasive coronary angiography, n (%)	683 (88%)	415 (87%)
Baseline medication		
Antiplatelet therapy, n (%)	775 (100%)	478 (100%)
Anticoagulants, n (%)	775 (100%)	478 (100%)
ACE-inhibitor/ARB, n (%)	775 (100%)	478 (100%)
Beta-blockers, n (%)	775 (100%)	478 (100%)
Nitrates or other antianginal drugs, n (%)	775 (100%)	478 (100%)
Calcium antagonists, n (%)	775 (100%)	478 (100%)
Diuretics, n (%)	775 (100%)	478 (100%)
Statins, n (%)	775 (100%)	478 (100%)
Insulin, n (%)	775 (100%)	478 (100%)
Other oral diabetic drugs, n (%)	775 (100%)	478 (100%)

Availability of data. The total proportion of missing data was 2% in the training set and 4% in the testing set. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blocker; BMI, body mass index; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; CT, computed tomography; ECG, electrocardiography; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; PCI, percutaneous coronary intervention; VHD, valvular heart disease.

Validation of machine learning-based risk stratification scores for patients with acute coronary syndrome treated with percutaneous coronary intervention

Mitchel A. Molenaar

Jasper L. Selder

Amand F. Schmidt

Folkert W. Asselbergs

Jelle D. Nieuwendijk

Brigitte van Dalfsen

Mark J. Schuuring

Berto J. Bouma

Steven A.J. Chamuleau

Niels J. Verouden

Abstract

Aims

This study aimed to validate the machine learning-based Global Registry of Acute Coronary Events (GRACE) 3.0 score and PRAISE (Prediction of Adverse Events following an Acute Coronary Syndrome) in patients with acute coronary syndrome (ACS) treated with percutaneous coronary intervention (PCI) for predicting mortality.

Methods and results

Data of consecutive patients with ACS treated with PCI in a tertiary centre in the Netherlands between 2014 and 2021 were used for external validation. The GRACE 3.0 score for predicting in-hospital mortality was evaluated in 2759 patients with non-ST-elevation acute coronary syndrome (NSTE-ACS) treated with PCI. The PRAISE score for predicting one-year mortality was evaluated in 4347 patients with ACS treated with PCI. Both risk scores were compared with the GRACE 2.0 score. The GRACE 3.0 score showed excellent discrimination [c-statistic 0.90 (95% CI 0.84, 0.94)] for predicting in-hospital mortality, with well-calibrated predictions (calibration-in-the large [CIL] -0.19 [95% CI -0.45, 0.07]). The PRAISE score demonstrated moderate discrimination [c-statistic 0.75 (95% CI 0.70, 0.80)] and overestimated the one-year risk of mortality [CIL -0.56 (95% CI -0.73, -0.39)]. Decision curve analysis demonstrated that the GRACE 3.0 score offered improved risk prediction compared with the GRACE 2.0 score, while the PRAISE score did not.

Conclusion

This study in ACS patients treated with PCI provides suggestive evidence that the GRACE 3.0 score effectively predicts in-hospital mortality beyond the GRACE 2.0 score. The PRAISE score demonstrated limited potential for predicting one-year mortality risk. Further external validation studies in larger cohorts including patients without PCI are warranted.

Introduction

Acute coronary syndrome (ACS) is defined as a range of conditions caused by sudden myocardial ischaemia. ACS affects more than 7 million people worldwide each year and approximately 8% of patients die within one year after admission^(1–4). To aid clinical decision making, risk assessment is performed in patients presenting with ACS^(5, 6). Numerous risk scores have been developed to assess the risk of in-hospital and longer term mortality, of which the Global Registry of Acute Coronary Events (GRACE) scoring system is recommended by the European Society of Cardiology (ESC) guideline^(5, 6). GRACE was a registry of patients presenting with ACS in the period 1999 to 2009 in a time before the widespread adoption of drug-eluting stents and contemporary medical therapy standards. Moreover, the generalizability of the GRACE 2.0 score remains limited, particularly in accurately predicting the probability of adverse events for the individual patient. These limitations indicated a need for new and more personalized risk stratification tools^(5, 6).

In recent years, machine learning-based risk scores have enabled the identification of more complex patterns compared with traditional regression methods^(7, 8). Machine learning has shown promising results for risk stratification in patients with ACS^(9, 10). For example, the machine learning-based GRACE version 3.0 score⁽¹¹⁾ improved the prediction of in-hospital mortality among patients with non-ST-elevation acute coronary syndrome (NSTE-ACS) compared with the GRACE version 2.0 score⁽¹¹⁾. The GRACE 3.0 score is calculated using nine variables of clinical presentation, which include the variables in the GRACE 2.0 score and sex. In external validation, the GRACE 3.0 score achieved a c-statistic of 0.91 (95% CI 0.89–0.92) in male patients and 0.87 (95% CI 0.84–0.89) in female patients⁽¹¹⁾. In addition, the PRAISE risk score⁽¹²⁾, a machine learning-based risk score, was trained on ACS patients treated with percutaneous coronary intervention (PCI) and integrates 25 clinical, anatomical, and procedural features. The PRAISE risk score achieved a c-statistic of 0.92 (95% CI 0.90–0.93) in an external validation cohort. This score has demonstrated accurate discrimination between patients who are likely to die within one year and those who are not⁽¹²⁾.

PCI is a widely utilized treatment in patients with ACS⁽¹³⁾. Validation of these machine learning-based tools is essential for selecting the most effective risk stratification tools, needed to better understand the patient's prognosis after PCI, identify high-risk individuals, and guide clinicians in their decisions. An independent validation study for the GRACE 3.0 score and PRAISE score in patients with ACS treated with PCI at a Dutch tertiary centre is lacking. Therefore, the aim of this study was to validate the GRACE 3.0 score and PRAISE in patients with ACS treated with PCI for predicting mortality.

Methods

Study design and patient population

In this retrospective cohort study, patients diagnosed by the treating physician with ACS were consecutively selected from electronic health records of the Amsterdam University Medical Cen-

ter, location VUmc, between 2015 and 2021. Patients were eligible if they were 18 years or older and underwent a PCI procedure. Patients were managed according to the ESC guidelines^(13–16). This study complies with the principles in the Declaration of Helsinki and received approval by the local human ethical review board. The study met the criteria for a waiver of the informed consent requirements.

Outcome

The primary outcomes were in-hospital mortality and one-year mortality.

Data collection

Baseline, treatment, and mortality data of patients who underwent PCI for ACS were collected from pseudonymized electronic health records and stored in a registry. Mortality data was verified using national registry data. The following variables were collected for each patient for calculation of the GRACE 2.0 score: age, heart rate, systolic blood pressure, Killip class, creatinine concentration, cardiac arrest, presence of ST-segment deviation, and troponin elevation. Calculation of the GRACE 3.0 score required the same variables, with addition of sex as variable. For calculation of the PRAISE score, 16 clinical variables (age, sex, diabetes, hypertension, hyperlipidemia, peripheral artery disease, estimated glomerular filtration rate [eGFR] using the Modification of Diet in Renal Disease [MDRD] equation⁽¹⁷⁾, previous myocardial infarction, previous PCI, previous coronary artery bypass grafting [CABG], previous stroke, previous bleeding, malignancy, STEMI presentation, haemoglobin, and left ventricular ejection fraction [LVEF]) were collected. The PRAISE score was calculated based on the collection of two procedural variables (vascular access and PCI with drug-eluting stent), two angiographic variables (multivessel disease and complete revascularization), and five therapeutic variables (treatment with β blockers, angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers, statins, oral anticoagulation, and proton-pump inhibitors). The definitions of these variables are shown in Supplementary Table 1 and were in accordance with the definitions used in the studies that developed the machine learning-based risk scores^(11, 12, 18). The GRACE 2.0 (in-hospital and one-year mortality), GRACE 3.0 (in-hospital mortality), and PRAISE scores (one-year mortality) were calculated for each patient using the online web pages (https://www.outcomes-umassmed.org/grace/acs_risk_2/index.html⁽⁴⁾, <https://www.grace-3.com>^(11, 18) and <https://praise.hpc4ai.it>⁽¹²⁾).

Missing data

Missing values of variables in the risk scores were imputed by multiple imputation by chained equation (MICE) in accordance with Wenzl et al.⁽¹¹⁾ MICE is an iterative process that imputes missing values using modelling of the other variables in the dataset. In this study, each variable was imputed 20 times (20 iterations), and this process was repeated to create 20 imputed datasets⁽¹⁹⁾. Predictive mean matching, proportional odds, and polytomous logistic regression were applied for continuous data, ordered categorical data, and unordered categorical data, respectively. The

outcome variables in-hospital mortality, six-month mortality, and one-year mortality were used as independent variables in the imputation process. Complete case analysis was performed to evaluate the effect of imputation on the results.

Evaluation of risk scores

The GRACE 3.0 score was evaluated in patients with NSTE-ACS treated with PCI for predicting in-hospital mortality⁽¹¹⁾. The PRAISE score was evaluated in patients with ACS treated with PCI for predicting one-year mortality. In line with the PRAISE development cohort score, patients who died during hospitalization were excluded^(20, 21). Additional analyses were conducted to investigate the performance of the GRACE 3.0 score and PRAISE score in patients with ACS and NSTE-ACS treated with PCI, in both male and female subgroups.

The discriminative performance of the machine learning-based risk scores was evaluated using concordance statistic (c-statistic)⁽²²⁾. The discriminative ability was classified as poor (c-statistic < 0.70), moderate (c-statistic 0.70–0.80), good (c-statistic 0.80–0.90), or excellent (c-statistic ≥ 0.90)^(22, 23). Rubin's Rules were used to combine the c-statistics from the imputed datasets and obtain an estimate with a 95% confidence interval⁽²⁴⁾.

The calibration (agreement between predicted and actual observed risk), was assessed using calibration plots, calibration-in-the-large (CIL), and the calibration slope (CS)^(25–27). CIL compares the average predicted risk with the observed risk, which is 0.00 in a perfectly calibrated risk score. A CIL lower than 0.00 indicates overestimation and a CIL greater than 0.00 indicates underestimation. CS evaluates the spread of the predicted risks, which is equal to 1.00 in a perfectly calibrated risk score. A CS lower than 1.00 indicates that the predicted risks are too high for patients at high risk and too low for patients at low risk. A CS greater than 1.00 indicates that the predicted risks are too low for high-risk patients and too high for low-risk patients^(26, 28). The calibration of the GRACE 3.0 score is particularly clinically important around the 3% threshold, which guides the decision on early vs. a delayed invasive treatment in patients with ACS⁽²⁹⁾.

Decision curve analysis was performed to evaluate the clinical utility of the machine learning-based risk scores for identifying high-risk and low-risk ACS patients of mortality after PCI^(30, 31). In this analysis, the net benefit (a weighted combination of true and false positives, determined by a threshold probability) of the risk score is plotted against a range of threshold probabilities. The threshold probabilities can be interpreted as the percentage at which a clinician would opt for close monitoring⁽³²⁾. This could be interpreted as the percentage of patients who require close monitoring needed to protect one patient from mortality. For example, setting a threshold at 0.05 (5%) means that to protect one patient from mortality, 20 patients need close monitoring. Similarly, for a threshold probability at 0.1 (10%), ten patients need close monitoring to prevent one patient from mortality. The net benefit across all thresholds of the machine learning risk scores is compared with the GRACE 2.0 score and close monitoring in all patients ('Always act'). The risk score with the highest net benefit at a certain threshold probability has the best trade-off between true positives and false positives and is the most clinically useful at that specific threshold probability. The maximum net-benefit is equal to the incidence of mortality, which is

the case when all patients who will die are identified by the risk score without any false positives. Decision curve analysis combines both discrimination and calibration, which offers a valuable method to compare the performance of risk scores.

Statistical analysis

Descriptive statistics were presented as median with interquartile range (IQR) to facilitate comparison with the cohorts upon which the risk scores (GRACE 2.0⁽⁴⁾, GRACE 3.0⁽¹¹⁾, PRAISE⁽¹²⁾) were developed. Nominal or ordinal data were presented as numbers with percentages. Data processing and statistical analysis were performed in Python version 3.8. and R version 4.1.3. (R Core Team, Vienna, Austria). The c-statistics of the GRACE 3.0 and PRAISE scores were compared with the GRACE 2.0 score on significance using the DeLong's test for paired receiver operating characteristics (ROC) curves⁽³³⁾. Rubin's Rules were used to aggregate the P-values obtained from the imputed datasets into a single estimate. A P-value less than 0.05 was considered statistically significant.

This study followed the TRIPOD (transparent reporting of multivariable prediction model for individual prognosis or diagnosis)⁽³⁴⁾ statement and met all CODE-EHR minimum framework standards for the use of healthcare data for clinical research⁽³⁵⁾.

Results

Study population

A total of 10411 patients who underwent CAG or PCI were screened for inclusion. After excluding patients without ACS or PCI, 4471 patients were identified. Of these patients, 2759 patients (62%) presented with NSTE-ACS, in whom the GRACE 3.0 risk score was evaluated. A total of 124 patients with ACS treated with PCI (2.7%) died during hospitalization, which resulted in 4347 patients with ACS in whom the PRAISE risk score was evaluated. A total number of 4418 patients with ACS treated with PCI were included in this study. The flow chart is depicted in Figure 1. In 24% of the patients, the data were not complete to calculate the GRACE 2.0 score. A total of 6.3% of the data were missing, as presented in Supplementary Table 2. Ejection fraction had the highest rate of missing data, which was missing in 61% of the patients. No data were missing for the outcomes in-hospital and one-year mortality.

A total of 3178 (72%) of patients in this study were male, and the median age was 66 years. The baseline characteristics are presented in Table 1. There were differences in the patient characteristics with the original GRACE 2.0⁽⁴⁾, GRACE 3.0⁽¹¹⁾, and PRAISE⁽¹²⁾ cohort, as shown in Supplementary Table 3. Patients in the GRACE 3.0 evaluation cohort had higher rates of cardiac arrest at admission and cardiogenic shock (Killip IV class) at presentation compared with patients in the GRACE 2.0 and 3.0 development cohort. PCI was performed in 35% of the females and 46% of the males in the original GRACE 3.0 cohort. The PRAISE development cohort had a higher prevalence of dyslipidemia, STEMI, and a higher median eGFR and ejection fraction, compared with patients in this study. All patients in the original PRAISE cohort underwent PCI.

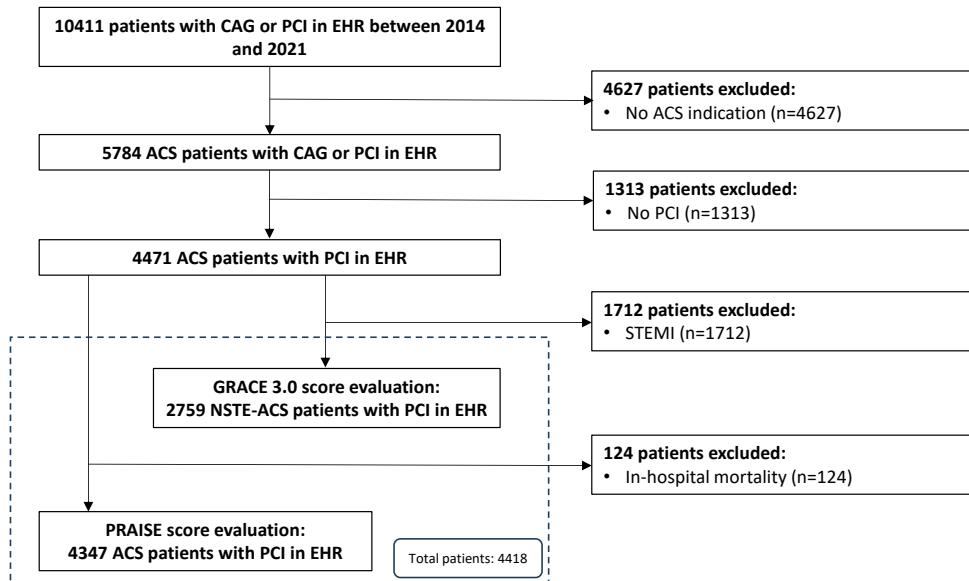


Figure 1: Inclusion flowchart. ACS = acute coronary syndrome; CAG = coronary angiography; EHR = electronic health records; NSTE-ACS = with non-ST-elevation acute coronary syndrome; PCI = percutaneous coronary intervention.

Mortality

In the GRACE 3.0 evaluation cohort 71 patients (2.5%) died in the hospital, and 156 patients (5.6%) died within one year after admission, as shown in Table 1. In the PRAISE evaluation cohort, 145 patients (3.3%) died within one year after admission.

Table 1: Baseline characteristics and adverse outcomes

Characteristics	PRAISE validation cohort (ACS, n=4347)	GRACE 3.0 validation cohort (NSTE-ACS, n=2759)
Age (years), median [Q1,Q3]	66.00 [57.00, 74.00]	68.00 [58.00, 75.00]
Male, n (%)	3127 (72.7)	1982 (72.6)
BMI (kg/m ²), median [Q1,Q3]	26.50 [24.21, 29.39]	26.60 [24.30, 29.70]
Risk factors		
Hypertension, n (%)	2234 (53.5)	1514 (57.3)
Diabetes, n (%)	910 (21.3)	672 (24.8)
Dyslipidemia, n (%)	1252 (30.1)	857 (32.6)
Current or former smoker, n (%)	1589 (38.1)	804 (30.4)
Medical history		
Myocardial infarction, n (%)	719 (16.5)	571 (20.7)
Percutaneous coronary intervention, n (%)	834 (19.2)	651 (23.6)
Coronary artery bypass graft, n (%)	220 (5.1)	196 (7.1)
Peripheral artery disease, n (%)	157 (3.9)	124 (4.8)
Stroke or transient ischaemic attack, n (%)	178 (4.1)	138 (5.0)
Bleedings, n (%)	63 (1.4)	46 (1.7)
Clinical presentation		
STEMI, n (%)	1659 (38.2)	0

Continued on next page

Characteristics	PRAISE validation cohort (ACS, n=4347)	GRACE 3.0 validation cohort (NSTE-ACS, n=2759)
Unstable angina or non-STEMI, n (%)	2688 (61.8)	2759 (100)
Haemoglobin at admission (mg/dL), median [Q1,Q3]	13.86 [12.57, 14.82]	13.70 [12.25, 14.82]
Heart rate (bpm), median [Q1,Q3]	70.00 [60.00, 81.00]	69.00 [60.00, 80.00]
Systolic blood pressure (mmHg), median [Q1,Q3]	127.00 [112.00, 143.00]	130.00 [114.00, 146.00]
Creatinine (mg/dL), median [Q1,Q3]	0.93 [0.79, 1.11]	0.95 [0.81, 1.14]
ST-segment deviation, n (%)	2396 (57.7)	768 (30.0)
Abnormal cardiac enzymes, n (%)	2559 (60.5)	1184 (44.6)
Cardiac arrest at admission, n (%)	169 (4.0)	107 (4.0)
Killip class, n (%)		
I	3293 (81.0)	2090 (81.6)
II	630 (15.5)	354 (13.8)
III	114 (2.8)	87 (3.4)
IV	28 (0.7)	31 (1.2)
Ejection fraction (%), median [Q1,Q3]	47.00 [35.00, 62.00]	47.00 [35.00, 62.00]
eGFR (mL/min/1.73 m ²), median [Q1,Q3]	63.44 [50.57, 80.76]	61.69 [48.96, 77.58]
Anatomy and procedural data		
Multivessel disease, n (%)	1981 (45.6)	1395 (50.6)
Percutaneous coronary intervention with DES implantation, n (%)	4347 (100.0)	2759 (100.0)
Complete revascularization, n (%)	3309 (76.1)	2142 (77.6)
Intervention		
Percutaneous coronary intervention, n (%)	4347 (100)	2759 (100)
Medical therapy at discharge		
Antiplatelet therapy, n (%)	3662 (84.2)	2345 (85.0)
Beta blockers, n (%)	2229 (51.3)	1214 (44.0)
ACE-inhibitor/ARB, n (%)	1870 (43.0)	1026 (37.2)
Statins, n (%)	2527 (58.1)	1270 (46.0)
Proton-pump inhibitor, n (%)	2190 (50.4)	1161 (42.1)
Oral anticoagulation, n (%)	466 (10.7)	309 (11.2)
Adverse outcomes		
In-hospital mortality, n (%)	0 (0)	71 (2.5)
One-year mortality, n (%)	145 (3.3)	156 (5.6)

Values are n (%), or median [interquartile range]. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blockers; BMI, body mass index; DES, drug-eluting stent; eGFR, estimated glomerular filtration rate [using the Modification of Diet in Renal Disease (MDRD) equation]; NSTE-ACS, non-ST-elevation acute coronary syndrome; STEMI, ST-elevation myocardial infarction.

Performance: GRACE 3.0 score

The c-statistic of the GRACE 3.0 score for predicting in-hospital mortality was 0.90 (95% CI 0.84, 0.94), which exceeded that of the GRACE 2.0 score [0.86 (95% CI 0.80, 0.91), P = 0.002]. The ROC curves are shown in Figure 2. While the GRACE 2.0 score [CIL -0.31 (95% CI -0.56, -0.06)] showed, on average, a slight overestimation of the risk of mortality, GRACE 3.0 score was well-calibrated [CIL -0.19 (95% CI -0.45, 0.07), Figure 3].

The CS of the GRACE 3.0 score [CS 0.96 (95% CI 0.81, 1.12)] and GRACE 2.0 score [CS 1.08 (95% CI 0.90, 1.25)] were nearly perfect, which indicates that the risk scores are well-calibrated at extreme high and low predictions. The calibration plots are shown in Figure 3. Due to missing data, 1868 patients with NSTE-ACS treated with PCI were included in complete case analysis. Complete case analysis is shown in Supplementary Table 4. In complete case analysis, 61 patients

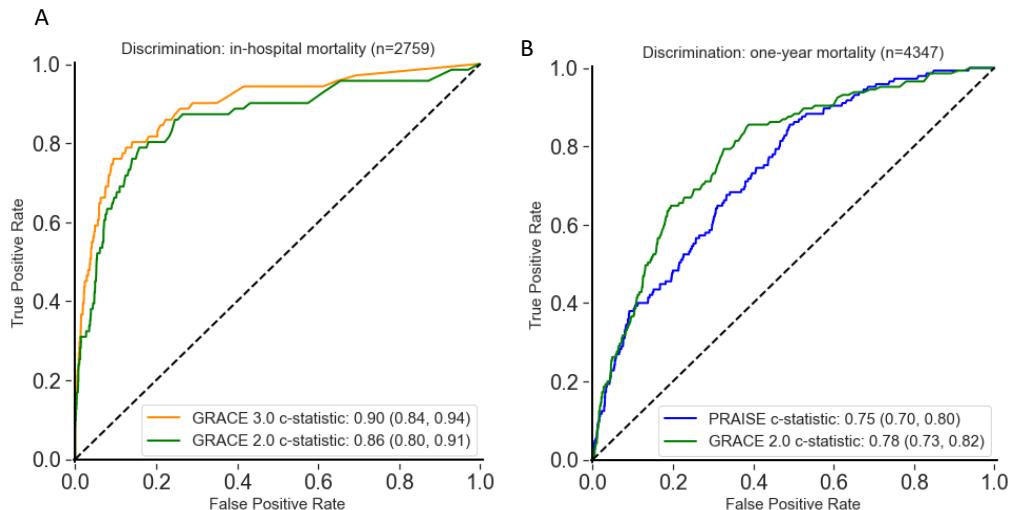


Figure 2: Discriminative performance of the GRACE 3.0 risk score, PRAISE risk score, and GRACE 2.0 risk score for predicting mortality. Legend: The ROC curves of the GRACE 3.0 and GRACE 2.0 scores for predicting in-hospital mortality (Plot A), as well as the PRAISE score and GRACE 2.0 score for predicting one-year mortality (Plot B), are depicted. The c-statistics are reported with a 95% confidence interval. GRACE = The Global Registry of Acute Coronary Events; PRAISE = Prediction of Adverse Events following an Acute Coronary Syndrome.

with NSTE-ACS treated with PCI (3%) died in the hospital. Complete case analysis yielded results similar to the analysis including all patients, except for the GRACE 2.0 [CIL -0.26 (95% CI -0.54, 0.01)] calibration, which was slightly better compared with the analysis of all patients.

Decision curve analysis showed that the GRACE 3.0 score was more effective in selecting patients at high- and low-risk of mortality for decision thresholds between 0% and 30%, compared with monitoring all patients closely and the GRACE 2.0 score. The decision curve analysis is shown in Figure 4.

In additional analyses, the c-statistic of the GRACE 3.0 risk score in male patients with NSTE-ACS treated with PCI ($n = 1982$) was 0.89 (95% CI 0.83, 0.95), and 0.93 (95% CI 0.79, 0.98) in female patients with NSTE-ACS treated with PCI ($n = 748$). These results are shown in Supplementary Table 5.

Performance: PRAISE score

The c-statistic of the PRAISE score for predicting one-year mortality was 0.75 (95% CI 0.70, 0.80), which was lower but not significantly different from that of the GRACE 2.0 score [0.78 (95% CI 0.73, 0.82), $P = 0.2$], as shown in Figure 2. The PRAISE score [CIL -0.56 (95% CI -0.73, -0.39)] showed near-perfect calibration for low mortality probabilities (0–5%). For predicted probabilities beyond 5%, the PRAISE score overestimated the risk of mortality. The GRACE 2.0 risk score overestimated the risk of mortality across the entire range of predicted mortality probabilities [CIL -1.12 (95% CI -1.30, -0.95)].

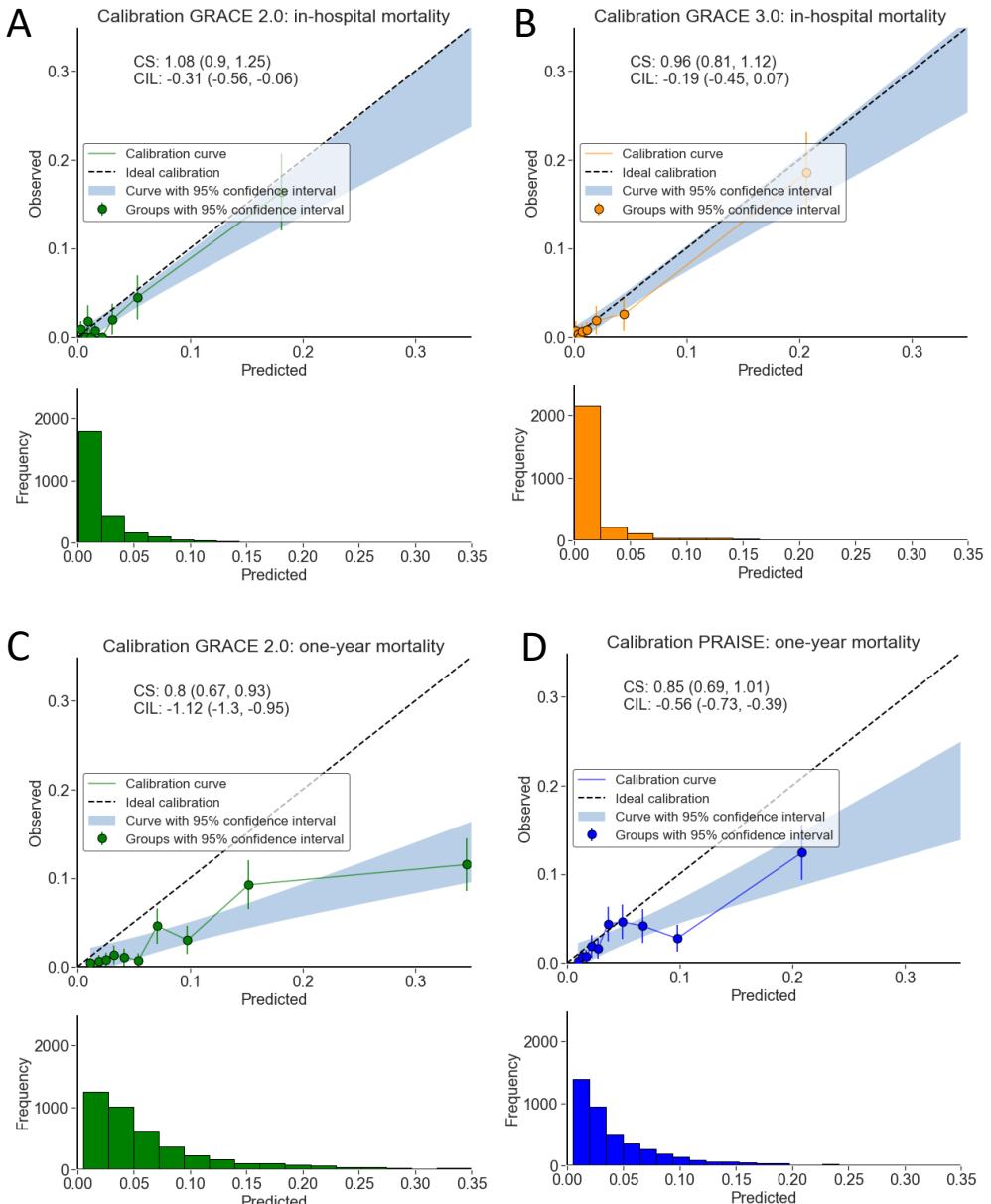


Figure 3: Calibration plots of the GRACE 3.0 risk score, PRAISE risk score, and GRACE 2.0 risk score. Legend: The calibration (agreement between predicted and actual observed risk) and frequency (number of patients falling within each predicted risk category) plots are depicted for the GRACE 2.0 (in-hospital mortality, Plot A), GRACE 3.0 (in-hospital mortality, Plot B), GRACE 2.0 (one-year mortality, Plot C), and PRAISE (one-year mortality, Plot D). Numbers are reported with 95% confidence intervals. CIL = calibration-in-the-large; CS = calibration slope.

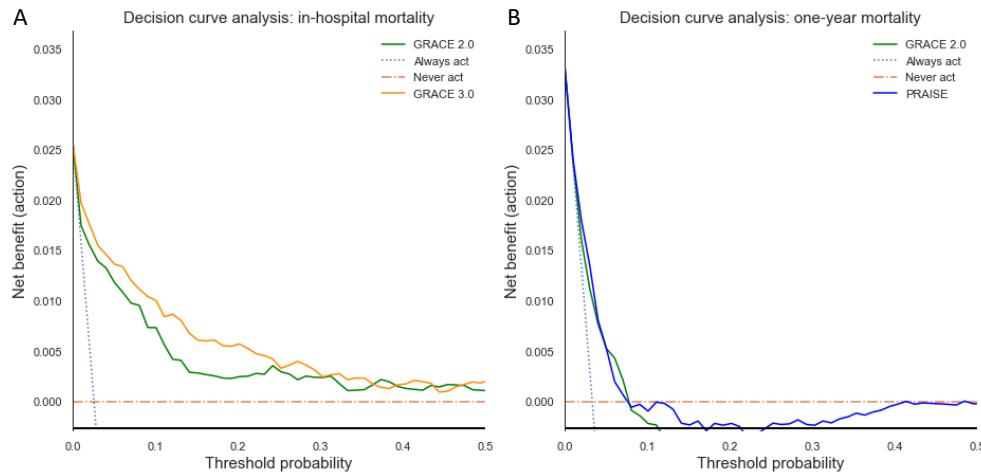


Figure 4: Decision curve analysis for the GRACE 3.0 risk score, GRACE 2.0 risk score, and the PRAISE risk score. Decision curve analysis of the GRACE 3.0 and GRACE 2.0 (Plot A), and PRAISE and GRACE 2.0 (Plot B). The y-axis represents the net benefit (a weighted combination of true and false positives, determined by the threshold probability) and the x-axis represents the threshold probability. A higher net benefit at a certain threshold probability indicates that the risk score is more clinically useful at that threshold probability for predicting mortality. Plot A demonstrates that the GRACE 3.0 risk score was more effective in selecting patients who require close monitoring compared with monitoring all patients closely ('Always act') and compared with the GRACE 2.0 risk score. Plot B demonstrates that the PRAISE score did not provide a significant benefit over the GRACE 2.0 score across different threshold probabilities.

Decision curve analysis of the PRAISE score did not show any improvement in risk prediction compared with the GRACE 2.0 score, as depicted in Figure 4.

In complete case analysis only 1174 ACS patients treated with PCI (27% of the patients) were included. Of these patients, a total of 63 (5%) died within one year after presentation. The c-statistic of the PRAISE score was 0.79 (95% CI 0.71, 0.85) and the model was well-calibrated [CIL 0.05 (-0.21, 0.32)].

In additional analyses, the c-statistic of the PRAISE risk score in male ACS patients treated with PCI ($n = 3127$) was 0.77 (95% CI 0.71, 0.82), and 0.71 (95% CI 0.62, 0.80) in female ACS patients treated with PCI ($n = 1177$). These results are shown in Supplementary Table 6.

Discussion

In this study, two machine learning-based risk scores (GRACE 3.0 and PRAISE) were validated in a population of patients with ACS who were treated with PCI in a tertiary centre. The GRACE 3.0 risk score, evaluated in 2759 patients with NSTEMI-ACS treated with PCI, showed excellent discriminative performance for predicting in-hospital mortality. The GRACE 3.0 risk score was more clinically useful as a risk prediction tool compared with the GRACE 2.0 score, as shown by the decision curve analysis. The PRAISE score, evaluated in 4347 ACS patients treated with PCI, showed moderate discrimination for predicting one-year mortality. The PRAISE score overestimated the risk of one-year mortality for patients with ACS treated with PCI with a predicted risk greater

than 5% and did not provide a significant benefit over the GRACE 2.0 score.

The GRACE 3.0 score was developed using prospective data from 386,591 patients with NSTEMI in England, Wales, and Northern Ireland. Wenzl et al. demonstrated comparable discriminatory performance to that observed in this study during validation in cohorts of external centres^(11, 36). The high discriminative ability can be explained by the machine learning model (XGBoost) that uses multiple decision trees to achieve optimal classification. In addition to the GRACE 2.0 score, the GRACE 3.0 score incorporates sex by utilizing separate machine learning models for male and female patients, which provides a more personalized sex-specific output. Previous studies have shown that the GRACE 2.0 score has suboptimal agreement between the predicted and observed risk of in-hospital mortality and one-year mortality^(6, 11, 37). The results of our study provide suggestive evidence that the GRACE 3.0 score could be a valuable tool for clinicians in assessing in-hospital mortality risk.

The PRAISE score was developed and validated by Ascenso et al.⁽¹²⁾, who found a high discriminatory performance [c-statistic 0.92 (95% CI 0.90, 0.93)] for this score in an external cohort. Differences in the patient population between our study and the study by Ascenso et al. may explain the limited ability of the risk score to generalize to a new population. In particular, the median eGFR and ejection fraction were higher in the patients of the study of Ascenso et al., which are important predictive factors of mortality in the PRAISE score⁽¹²⁾. Shi et al.⁽³⁸⁾ validated the PRAISE score in an Asian population (6412 ACS patients treated with PCI) and demonstrated that the PRAISE had a slightly greater net benefit compared with the GRACE 2.0 score. In line with Shi et al., the results of our study suggest that retraining of the PRAISE model on our dataset is required to improve the prediction of mortality for this risk score.

The PRAISE score was calculated using 25 variables, including anatomical and procedural data. Anatomical and procedural data have been previously incorporated into mortality prediction scores⁽³⁹⁾, such as vascular access site⁽⁴⁰⁾, number of diseased coronary arteries⁽⁴¹⁻⁴³⁾, and location of significant lesions⁽⁴³⁾. The impact of anatomical and procedural data in a machine learning model for risk prediction has been investigated in only a limited number of studies. Zack et al.⁽⁴⁴⁾ demonstrated that their model outperformed logistic regression (c-statistic 0.88 vs. c-statistic 0.81) for predicting long term (180-day) mortality using 410 variables, including angiographic and interventional details. In addition, Mori et al.⁽⁴⁵⁾ trained a XGBoost model on angiographic data of 378 572 patients treated with CABG to predict multiple outcomes and demonstrated improved risk stratification compared with machine learning models trained on clinical data alone. These findings suggest that further research is needed to establish a set of important anatomical and procedural features to further explore the potential of machine learning for optimizing risk stratification in ACS patients treated non-conservatively. It is important to note that calculating a risk score involving a large number of variables via calculators can increase clinicians' workload. Automatic calculation and integration in electronic health records can help clinicians predict the risk for individual patients.

The number of clinical prediction models is growing exponentially. However, only a fraction of these models is validated in an external cohort. Wessler et al. demonstrated that only 42%

of the cardiovascular clinical prediction models were validated in an external cohort⁽⁴⁶⁾. External validation is needed to determine the generalizability of a model, especially in models that are trained on small datasets. External validation is a prerequisite before implementing a model in clinical practice⁽⁴⁷⁾. In the context of patients with ACS, a risk model can have implications for treatment decisions by, for example, selecting the right patients who may require close monitoring, aggressive management of risk factors, extended hospital stay, additional interventions, or for prioritizing treatment. Therefore, validation of these machine learning models in diverse clinical settings is important, to which our study contributes.

Several remarks can be made about this study. First, patients were retrospectively included in our study, which resulted in missing data. Complete case analysis and all-case analysis did not show significant differences in the performance of the GRACE 3.0 score. The PRAISE score overestimated the risk of one-year mortality in all-case analysis, while the risk score demonstrated nearly perfect calibration in complete case analysis. These findings may be explained by the substantial number of patients who had missing data, with variables missing in up to 61% of the patients. Excluding patients with missing data from analysis resulted in a different risk distribution compared with the entire population, with higher rates of mortality (5.0% vs. 3.3%). Second, the GRACE 3.0 score appeared to be more clinically useful compared with the GRACE 2.0 score. The GRACE 3.0 score was originally developed on a population of NSTE-ACS patients in which 36% females and 64% males underwent PCI. It was not specifically tailored for patients who underwent PCI, as in our study. It is possible that we have primarily selected patients with a lower clinician-assessed risk of poor outcomes compared with the original GRACE 3.0 cohort of Wenzl et al. due to the risk-treatment paradox^(6, 48, 49) in which clinicians might be hesitant to perform an invasive procedure in patients with a high risk of adverse outcomes. Further assessment is needed in an untreated ACS population to confirm the generalizability of the GRACE 3.0 score for risk stratification. Third, the available data limited our ability to investigate the performance of the PRAISE risk score (recurrent ACS, major bleeding) beyond all-cause mortality and to evaluate other risk scores (e.g. PARIS and PRECISE-DAPT risk scores). Fourth, the combination of the single-centre design and the low number of in-hospital mortality events necessitates larger datasets to conclusively assess the performance of the evaluated risk scores. Fifth, a strength of the study is that the machine learning-based risk scores were compared with the GRACE 2.0 score, which is recommended by the ESC guidelines for risk assessment in patients with ACS⁽²⁹⁾.

Conclusions

In conclusion, this study in ACS patients treated with PCI provides suggestive evidence that the GRACE 3.0 score effectively predicts in-hospital mortality beyond the GRACE 2.0 score. In our dataset, the PRAISE score showed limited potential for predicting one-year mortality and did not enhance clinical decision making compared with the GRACE 2.0 score. Further external validation of GRACE 3.0 and PRAISE in larger prospective multi-centre patient cohorts including patients without PCI is warranted.

References

- [1] Reed GW, Rossi JE, Cannon CP. Acute myocardial infarction. *The Lancet* 2017;389:197–210.
- [2] Steg PG, Goldberg RJ, Gore JM, Fox KAA, Eagle KA, Flather MD, et al. Baseline characteristics, management practices, and in-hospital outcomes of patients hospitalized with acute coronary syndromes in the Global Registry of Acute Coronary Events (GRACE). *The American Journal of Cardiology* 2002;90:358–363.
- [3] Bhatt DL, Lopes RD, Harrington RA. Diagnosis and Treatment of Acute Coronary Syndromes: A Review. *JAMA* 2022;327:662–675.
- [4] Fox KAA, Fitzgerald G, Puymirat E, Huang W, Carruthers K, Simon T, et al. Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ Open* 2014;4:e004425.
- [5] Ono M, Kawashima H, Hara H, Gamal A, Wang R, Gao C, et al. External validation of the GRACE risk score 2.0 in the contemporary all-comers GLOBAL LEADERS trial. *Catheter Cardiovasc Interv* 2021;98:E513–E522.
- [6] Sanger NMR van der, Azzahhafi J, Yin DRPPCP, Peper J, Rayhi S, Walhout RJ, et al. External validation of the GRACE risk score and the risk–treatment paradox in patients with acute coronary syndrome. *Open Heart* 2022;9:e001984.
- [7] Oikonomou EK, Williams MC, Kotanidis CP, Desai MY, Marwan M, Antonopoulos AS, et al. A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography. *Eur Heart J* 2019;40:3529–3543.
- [8] Molenaar MA, Bouma BJ, Asselbergs FW, Verouden NJ, Selder JL, Chamuleau SAJ, et al. Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome. *Eur Heart J Digit Health* 2024;5:170–182.
- [9] Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, et al. Current State and Future Perspectives of Artificial Intelligence for Automated Coronary Angiography Imaging Analysis in Patients with Ischemic Heart Disease. *Curr Cardiol Rep* 2022;24:365–376.
- [10] Gill SK, Karwath A, Uh H-W, Cardoso VR, Gu Z, Barsky A, et al. Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *Eur Heart J* 2023;44:713–725.
- [11] Wenzl FA, Krämer S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *The Lancet* 2022;400:744–756.
- [12] D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *The Lancet* 2021;397:199–207.
- [13] Steg P, James SK, Atar D, Badano LP, Lundqvist CB, Borger MA, et al. ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). *Eur Heart J* 2012;33:2569–2619.
- [14] Windecker S, Kohl P, Alfonso F, Collet J-P, Cremer J, Falk V, et al. 2014 ESC/EACTS Guidelines on myocardial revascularization: The Task Force on Myocardial Revascularization of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS) Developed with the special contribution of the European Association of Percutaneous Cardiovascular Interventions (EAPCI). *Eur Heart J* 2014;35:2541–2619.
- [15] Roffi M, Patrono C, Collet J-P, Mueller C, Valgimigli M, Andreotti F, et al. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2016;37:267–315.
- [16] Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2018;39:119–177.
- [17] Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999;130:461–470.
- [18] Wenzl FA, Lüscher TF. Application of a sex-specific GRACE score in practice – Authors' reply. *The Lancet* 2023;401:23.
- [19] Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
- [20] D'Ascenzo F, Biondi-Zocca G, Moretti C, Bollati M, Omedè P, Scitto F, et al. TIMI, GRACE and alternative risk scores in acute coronary syndromes: a meta-analysis of 40 derivation studies on 216,552 patients and of 42 validation studies on 31,625 patients. *Contemp Clin Trials* 2012;33:507–514.
- [21] De Filippo O, D'Ascenzo F, Raposeiras-Roubin S, Abu-Assi E, Peyracchia M, Bocchino PP, et al. P2y12 inhibitors in acute coronary syndrome patients with renal dysfunction: an analysis from the RENAMI and Bleeding MACS projects. *Eur Heart J Cardiovasc Pharmacother* 2020;6:31–42.
- [22] Assessing the Fit of the Model. Applied logistic regression. Hoboken, NJ: John Wiley & sons, Ltd; 2000. p.143–202.
- [23] Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation* 2010;121:1768–1777.
- [24] RUBIN DB. Inference and missing data. *Biometrika* 1976;63:581–592.
- [25] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–1931.

- [26] Van Calster B, McLernon DJ, Smeden M van, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17:230.
- [27] Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ* 2024;384:e074820.
- [28] Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;45:562–565.
- [29] Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, et al. 2023 ESC guidelines for the management of acute coronary syndromes: developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J* 2023;44:3720–3826.
- [30] Vickers AJ, Calster B van, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019;3:18.
- [31] Vickers AJ, Calster BV, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6.
- [32] Vickers AJ, Van Claster B, Wynants L, Steyerberg EW. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. *Diagn Progn Res* 2023;7:11.
- [33] 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.
- [34] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13:1.
- [35] Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *Eur Heart J* 2022;43:3578–3588.
- [36] Wenzl FA, Bruno F, Kofoed KF, Raebler L, Roffi M, Stellios K, et al. Validation of the GRACE 3.0 score and redefinition of the risk threshold for early invasive treatment in non-ST-segment elevation acute coronary syndromes: a modelling study from five countries. *Eur Heart J* 2023;44:e025812.
- [37] Hung J, Roos A, Kadesjö E, McAllister DA, Kimenai DM, Shah ASV, et al. Performance of the GRACE 2.0 score in patients with type 1 and type 2 myocardial infarction. *Eur Heart J* 2020;42:2552–2561.
- [38] Shi B, Wang H, Liu J, Cai Z, Song C, Yin D, et al. Prognostic Value of Machine-Learning-Based PRAISE Score for Ischemic and Bleeding Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention. *J Am Heart Assoc* 2023;12:e025812.
- [39] Hizoh I, Domokos D, Banhegyi G, Becker D, Merkely B, Ruzsa Z. Mortality prediction algorithms for patients undergoing primary percutaneous coronary intervention. *J Thorac Dis* 2020;12:1706–1720.
- [40] Hizoh I, Gulyas Z, Domokos D, Banhegyi G, Majoros Z, Major L, et al. A novel risk model including vascular access site for predicting 30-day mortality after primary PCI: The ALPHA score. *Cardiovasc Revasc Med* 2017;18:33–39.
- [41] De Luca G, Suryapranata H, Hof AWJ van't, Boer M-J de, Hoornstje JCA, Dambrink J-HE, et al. Prognostic Assessment of Patients With Acute Myocardial Infarction Treated With Primary Angioplasty. *Circulation* 2004;109:2737–2743.
- [42] Halkin A, Singh M, Nikolsky E, Grines CL, Tcheng JE, Garcia E, et al. Prediction of mortality after primary percutaneous coronary intervention for acute myocardial infarction: the CADILLAC risk score. *J Am Coll Cardiol* 2005;45:1397–1405.
- [43] Mulder M de, Gitt A, Domburg R van, Hochadel M, Seabra-Gomes R, Serruys PW, et al. EuroHeart score for the evaluation of in-hospital mortality in patients undergoing percutaneous coronary intervention. *Eur Heart J* 2011;32:1398–1408.
- [44] Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar -Sinai Yoav, Widmer RJ, et al. Leveraging Machine Learning Techniques to Forecast Patient Prognosis After Percutaneous Coronary Intervention. *JACC Cardiovasc Interv* 2019;12:1304–1311.
- [45] Mori M, Durant TJS, Huang C, Mortazavi BJ, Coppi A, Jean RA, et al. Toward Dynamic Risk Prediction of Outcomes After Coronary Artery Bypass Graft: Improving Risk Prediction With Intraoperative Events Using Gradient Boosting. *Circ Cardiovasc Qual Outcomes* 2021;14:e007363.
- [46] Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External Validations of Cardiovascular Clinical Prediction Models: A Large-Scale Review of the Literature. *Circ Cardiovasc Qual Outcomes* 2021;14:e007858.
- [47] Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derkzen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826–832.
- [48] Hall M, Bebb OJ, Dondo TB, Yan AT, Goodman SG, Bueno H, et al. Guideline-indicated treatments and diagnostics, GRACE risk score, and survival for non-ST elevation myocardial infarction. *Eur Heart J* 2018;39:3798–3806.
- [49] Saar A, Marandi T, Ainla T, Fischer K, Blöndal M, Eha J. The risk-treatment paradox in non-ST-elevation myocardial infarction patients according to their estimated GRACE risk. *Int J Cardiol* 2018;272:26–32.

Supplementary material

Supplementary Table 1: Definition of variables

Variable	Definition
Killip class	Class 1: No heart failure. No clinical signs of cardiac decompensation. Class 2: Heart failure. Diagnostic criteria include rales, S3 gallop, and venous hypertension. Class 3: Frank pulmonary edema. Class 4: Cardiogenic shock. Signs include hypotension (systolic pressure of 90 mm Hg or less) and evidence of peripheral vasoconstriction such as oliguria, cyanosis, and diaphoresis. ¹
Troponin elevation	> 14.0 ng/L ² .
Hypertension	History of high blood pressure diagnosed or treated by a physician.
Diabetes mellitus	History of diabetes mellitus diagnosed or treated by a physician, or being in treatment with hypoglycemic drugs.
Hyperlipidaemia	History of hyperlipidemia diagnosed or treated by a physician.
Peripheral artery disease	History of peripheral artery disease diagnosed or treated by a physician.
estimated glomerular filtration rate (eGFR)	Calculated using the Modification of Diet in Renal Disease Study formula ³ .
Myocardial infarction	History of myocardial infarction diagnosed or treated by a physician.
Previous stroke or transient ischemic attack	History of stroke or transient ischemic attack diagnosed or treated by a physician.
Previous bleeding	History of bleeding diagnosed or treated by a physician.
Left ventricular ejection fraction (LVEF)	Assessed by 2D transthoracic echocardiography and computed according to Simpson's method of disks [(left ventricular end diastolic volume – left ventricular end systolic volume)/ left ventricular end diastolic volume].
Multivessel disease	Presence of multivessel disease at the time of the current intervention. Defined as at least a 70% stenosis in 2 or more native coronary arteries (RCA / LAD / RCX, or in the first-order side branch with a diameter of at least 1.5mm).
Complete revascularization	No residual significant stenosis (\geq 70% diameter stenosis at angiographic evaluation) in any coronary vessel after PCI.

Values are n (%), or median [interquartile range]. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blockers; BMI, body mass index; DES, drug-eluting stent; eGFR, estimated glomerular filtration rate [using the Modification of Diet in Renal Disease (MDRD) equation]; NSTE-ACS, non-ST-elevation acute coronary syndrome; STEMI, ST-elevation myocardial infarction.

Supplementary Table 2: Missing Data

Characteristics	ACS (n=4418)	PRAISE (ACS, n=4347)	GRACE 3.0 (NSTE-ACS, n=2759)
Age (years), n (%)	0 (0)	0 (0)	0 (0)
Male, n (%)	44 (1.0)	43 (1.0)	29 (1.1)
BMI (kg/m ²), n (%)	1680 (38.0)	1660 (38.2)	806 (29.2)
Risk factors			
Hypertension, n (%)	174 (3.9)	172 (4.0)	116 (4.2)
Diabetes, n (%)	71 (1.6)	70 (1.6)	54 (2.0)
Dyslipidemia, n (%)	195 (4.4)	193 (4.4)	131 (4.7)
Current or former smoker, n (%)	178 (4.0)	176 (4.0)	118 (4.3)
Medical History			
Myocardial infarction, n (%)	0 (0)	0 (0)	0 (0)
Percutaneous coronary intervention, n (%)	0 (0)	0 (0)	0 (0)
Coronary artery bypass graft, n (%)	0 (0)	0 (0)	0 (0)
Peripheral artery disease, n (%)	285 (6.5)	282 (6.5)	197 (7.1)
Stroke or transient ischaemic attack, n (%)	0 (0)	0 (0)	0 (0)
Bleedings, n (%)	0 (0)	0 (0)	0 (0)
Clinical presentation			
STEMI, n (%)	0 (0)	0 (0)	0 (0)
Unstable angina or non-STEMI, n (%)	0 (0)	0 (0)	0 (0)
Haemoglobin at admission (mg/dL), n (%)	1211 (27.4)	1206 (27.7)	1118 (40.5)
Heart rate (bpm), n (%)	98 (2.2)	98 (2.3)	76 (2.8)
Systolic blood pressure (mmHg), n (%)	91 (2.1)	91 (2.1)	77 (2.8)
Creatinine (mg/dL), n (%)	776 (17.6)	769 (17.7)	705 (25.6)
ST-segment deviation, n (%)	197 (4.5)	194 (4.5)	197 (7.1)
Abnormal cardiac enzymes, n (%)	119 (2.7)	117 (2.7)	105 (3.8)
Cardiac arrest at admission, n (%)	81 (1.8)	80 (1.8)	60 (2.2)

Continued on next page

Characteristics	ACS (n=4418)	PRAISE (ACS, n=4347)	GRACE 3.0 (NSTE-ACS n=2759)
Killip class, n (%)	285 (6.5)	282 (6.5)	197 (7.1)
Ejection fraction (%), n (%)	2693 (61.0)	2670 (61.4)	1457 (52.8)
eGFR (ml/min/1.73 m ²), n (%)	776 (17.6)	769 (17.7)	705 (25.6)
Anatomy and procedural data			
Multivessel disease, n (%)	0 (0)	0 (0)	0 (0)
Percutaneous coronary intervention with DES implantation, n (%)	0 (0)	0 (0)	0 (0)
Complete revascularization, n (%)	0 (0)	0 (0)	0 (0)
Medical therapy at discharge			
Antiplatelet therapy, n (%)	0 (0)	0 (0)	0 (0)
Beta blockers, n (%)	0 (0)	0 (0)	0 (0)
ACE-inhibitor/ARB, n (%)	0 (0)	0 (0)	0 (0)
Statins, n (%)	0 (0)	0 (0)	0 (0)
Proton-pump inhibitor, n (%)	0 (0)	0 (0)	0 (0)
Oral anticoagulation, n (%)	0 (0)	0 (0)	0 (0)

Values are n (%). The total proportion of missing data was 5.7%. ACE=angiotensin-converting enzyme; ARB=angiotensin receptor blockers; BMI=body mass index; DES=drug-eluting stent; eGFR=estimated glomerular filtration rate (using the Modification of Diet in Renal Disease [MDRD] equation3); STEMI=ST-elevation myocardial infarction.

Supplementary Table 3: Baseline characteristics and adverse outcomes

Characteristics	GR2.0 Dev (n=32037) ⁴	GR3.0 Dev F (n=145738) ⁵	GR3.0 Dev M (n=254316) ⁵	PR Dev (n=19826) ⁶
Age (years), median [Q1, Q3]	66.6 [56, 76]	76 (66-84)	69 (58-79)	64 (54-73)
Male, n (%)	67	0	100	78
BMI (kg/m ²), median [Q1, Q3]	27 [24-30]	26.5 (23.0-30.9)	27.4 (24.6-30.8)	
Risk factors				
Hypertension, n (%)	64	59.2	52.7	55.9
Diabetes, n (%)	26	24.1 (type 2)	24.9 (type 2)	24.8
Dyslipidemia, n (%)	51	33.9	37.7	51.3
Current or former smoker, n (%)	58	17.6	24.1	
Medical History				
Myocardial infarction, n (%)	30		12.6	
Percutaneous coronary intervention, n (%)	19	8.8	14.3	12.8
Coronary artery bypass graft, n (%)	13	4.8	10.6	2.7
Peripheral artery disease, n (%)	9	4.3	5.6	5.6
Stroke or transient ischaemic attack, n (%)	8.5	11.1	9.2	5.8
Bleedings, n (%)				4.9
Clinical Presentation				
STEMI, n (%) or %	36	0	0	56.5
Unstable angina or non-STEMI, n (%) or %		100	100	43.8
Haemoglobin at admission (mg/dL), median [Q1,Q3]		12.8 (11.5-13.9)	14.0 (12.5-15.1)	14 [12,17]
Heart rate (bpm), median [Q1,Q3]	76 [65,90]	80 (69-95)	76 (65-90)	
Systolic blood pressure (mmHg), median [Q1,Q3]140 [120,160]		142 (123-161)	139 (122-157)	
Creatinine (mg/dL), median [Q1,Q3]	1.02 [0.90-1.25]	0.9 (0.7-1.2)	1.1 (0.9-1.3)	
ST-segment deviation, n (%) or %	53	24.7	25.2	
Abnormal cardiac enzymes, n (%) or %	52	90.1	90.1	
Cardiac arrest at admission, n (%) or %	1.9	0.6	1.2	
Killip class, n (%) or %				
I	85	74.5	80.4	
II	11	18.7	14.1	
III	3.6	7.2	5.0	
IV	0.8	0.5	0.6	
Ejection fraction (%), median [Q1,Q3]				55 [39,61]
eGFR (ml/min/1.73 m ²), median [Q1,Q3]		62.1 (43.7-81.1)	72.8 (53.4-89.0)	83 [49,117]
Anatomy and Procedural Data				
Multivessel disease, n (%) or %				58.2
Percutaneous coronary intervention with DES implantation, n (%) or %				59.6
Complete revascularization, n (%) or %				57.5
Intervention				
Percutaneous coronary intervention, n (%) or %	35	46	100	
Medical Therapy at Discharge				

Continued on next page

Characteristics	GR2.0 Dev (n=32037) ⁴	GR3.0 Dev F (n=145738) ⁵	GR3.0 Dev M (n=254316) ⁵	PR Dev (n=19826) ⁶
Antiplatelet therapy, n (%) or %				
Beta blockers, n (%) or %	33 (at presentation)	33 (at presentation)		81.6
ACE-inhibitor/ARB, n (%) or %	42 (at presentation)	42 (at presentation)		81.6
Statins, n (%) or %	46 (at presentation)	46 (at presentation)		93.1
Proton-pump inhibitor, n (%) or %				34.7
Oral anticoagulation, n (%) or %				4.6
Adverse Outcomes				
In-hospital mortality, n (%) or %	4.0	2.5	1.7	
One-year mortality, n (%) or %	8	8.9	5.2	3.3

Values are n (%) or median [interquartile range]. The total proportion of missing data was 5.7%. ACE=angiotensin-converting enzyme; ARB=angiotensin receptor blockers; BMI=body mass index; DES=drug-eluting stent; DEV=development cohort; eGFR=estimated glomerular filtration rate (using the Modification of Diet in Renal Disease [MDRD] equation³); F=Female; GR=GRACE; M=Male; PR=PRAISE; STEMI=ST-elevation myocardial infarction.

Supplementary Table 4: Complete case-analysis

Model	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)	C-statistic (95% CI)	DeLong p-value
Inhospital Mortality				
Grace 2.0 (n=1868)	1.06 (0.86, 1.26)	-0.26 (-0.54, 0.01)	0.86 (0.80, 0.92)	<0.001
Grace 3.0 (n=1868)	0.98 (0.8, 1.16)	-0.16 (-0.45, 0.12)	0.90 (0.85, 0.95)	
One-year Mortality				
Grace 2.0 (n=1174)	0.7 (0.5, 0.91)	-0.99 (-1.27, -0.72)	0.76 (0.68, 0.82)	0.36
PRAISE (n=1174)	1.06 (0.8, 1.32)	0.05 (-0.21, 0.32)	0.79 (0.71, 0.85)	

CI=confidence interval

Supplementary Table 5: Additional analyses GRACE 3.0 risk score

Model	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)	C-statistic (95% CI)	DeLong p-value
NSTE-ACS Male				
Grace 2.0 (n=1982)	0.98 (0.78, 1.18)	-0.3 (-0.6, -0.01)	0.84 (0.77, 0.91)	0.002
Grace 3.0 (n=1982)	0.89 (0.71, 1.06)	-0.12 (-0.44, 0.19)	0.89 (0.83, 0.95)	
NSTE-ACS Female				
Grace 2.0 (n=748)	1.42 (0.99, 1.85)	-0.45 (-0.93, 0.03)	0.91 (0.77, 0.97)	0.18
Grace 3.0 (n=748)	1.34 (0.94, 1.75)	-0.4 (-0.89, 0.08)	0.93 (0.79, 0.98)	

CI=confidence interval

Supplementary Table 6: Additional analyses PRAISE risk score

Model	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)	C-statistic (95% CI)	DeLong p-value
ACS Male				
Grace 2.0 (n=3127)	0.71 (0.55, 0.86)	-1.15 (-1.36, -0.94)	0.76 (0.71, 0.81)	0.63
PRAISE (n=3127)	0.87 (0.69, 1.06)	-0.62 (-0.82, -0.41)	0.77 (0.71, 0.82)	
ACS Female				
Grace 2.0 (n=1177)	1 (0.74, 1.25)	-1.06 (-1.37, -0.75)	0.83 (0.74, 0.89)	0.005
PRAISE (n=1177)	0.78 (0.47, 1.09)	-0.48 (-0.78, -0.18)	0.71 (0.62, 0.80)	

CI=confidence interval

References Appendix

1. Killip T, Kimball JT. Treatment of myocardial infarction in a coronary care unit. A two year experience with 250 patients. *Am J Cardiol* 1967;20:457–464.
2. Wenzl FA, Lüscher TF. Application of a sex-specific GRACE score in practice – Authors' reply. *The Lancet* 2023;401:23.
3. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999;130:461–470.
4. Fox KAA, FitzGerald G, Puymirat E, Huang W, Carruthers K, Simon T, et al. Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ Open* 2014;4:e004425.
5. Wenzl FA, Kräler S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *The Lancet* 2022;400:744–756.
6. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *The Lancet* 2021;397:199–207.

Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease

Mitchel A. Molenaar

Jasper L. Selder

Johny Nicolas

Bimmer E. Claessen

Roxana Mehran

Javier Oliván Bescós

Mark J. Schuuring

Berto J. Bouma

Niels J. Verouden

Steven A.J. Chamuleau

Abstract

Purpose of Review

Artificial intelligence (AI) applications in (interventional) cardiology continue to emerge. This review summarizes the current state and future perspectives of AI for automated imaging analysis in invasive coronary angiography (ICA).

Recent Findings

Recently, 12 studies on AI for automated imaging analysis in ICA have been published. In these studies, machine learning (ML) models have been developed for frame selection, segmentation, lesion assessment, and functional assessment of coronary flow. These ML models have been developed on monocenter datasets (in range 31–14,509 patients) and showed moderate to good performance. However, only three ML models were externally validated.

Summary

Given the current pace of AI developments for the analysis of ICA, less-invasive, objective, and automated diagnosis of CAD can be expected in the near future. Further research on this technology in the catheterization laboratory may assist and improve treatment allocation, risk stratification, and cath lab logistics by integrating ICA analysis with other clinical characteristics.

Introduction

Artificial intelligence (AI) has an emerging role in healthcare in general, and the same holds for cardiology specifically, with numerous solutions in cardiac imaging modalities on image acquisition and reconstruction, diagnosis, and prognosis⁽¹⁾. For example, AI applications are now being utilized to accelerate acquisition and reduce reconstruction time of cardiac MRI, to automate disease classification in echocardiography, and to improve conventional risk prediction models based on coronary CT angiography features⁽²⁻⁵⁾. Despite growing applications in general cardiology, the role of AI in automated analysis of invasive coronary angiography (ICA) is less clear. ICA is an indispensable step in the diagnosis of coronary artery disease (CAD) in symptomatic patients⁽⁶⁾. This invasive imaging modality assesses the severity of stenoses by X-ray imaging of contrast-filled coronary arteries. In case of significant CAD, a multidisciplinary heart team decides on an appropriate treatment strategy, either conservative management or percutaneous or surgical revascularization. The heart team assessment is largely based on ICA assessment in combination with clinical parameters. Furthermore, percutaneous coronary interventions (PCIs) are guided by ICA for identification of target lesions; determining wiring, lesion preparation, and stenting strategies; and evaluation of procedural success based on residual stenosis, absence of significant dissection, and flow⁽⁶⁾.

After a general introduction of AI (as an application), we summarize the current state of AI for ICA imaging analysis and discuss its clinical implications for diagnosis, (real-time) treatment guidance, and risk stratification. We conclude this review with a discussion of its current limitations and future perspectives.

Artificial Intelligence: a Deeper Understanding

Artificial intelligence (AI) has become a collective term for applications that perform complex tasks that previously required human intelligence. Machine learning (ML), a subfield of AI, is performing complex tasks by learning from experience. Training of an ML algorithm creates an ML model, which represents what was learned by the ML algorithm to make predictions on new data. Most common ML applications in cardiac imaging can be broadly subdivided into two categories: supervised learning and unsupervised learning. In supervised learning, categorized data are used to classify unseen data. An example of supervised learning is the training of ML algorithms to predict a patient's response to certain treatment. In unsupervised learning, ML algorithms are trained to find patterns or conclusions through unlabeled training data. A well-known unsupervised learning method is clustering in which data/patients are grouped on similarity, for example, to identify distinct clinical subgroups of patients which may benefit from targeted therapy^(7, 8).

Deep learning (DL) is a subfield of ML in which multilayered neural networks are trained to learn a supervised or unsupervised task. A neuron is a mathematical function that provides an output based on the input. During training, weights of the neurons in a neural network are optimized to map the input(s) to a desired output. Feature selection is an important processing step to select relevant input variables before training an AI algorithm. The selection of features that

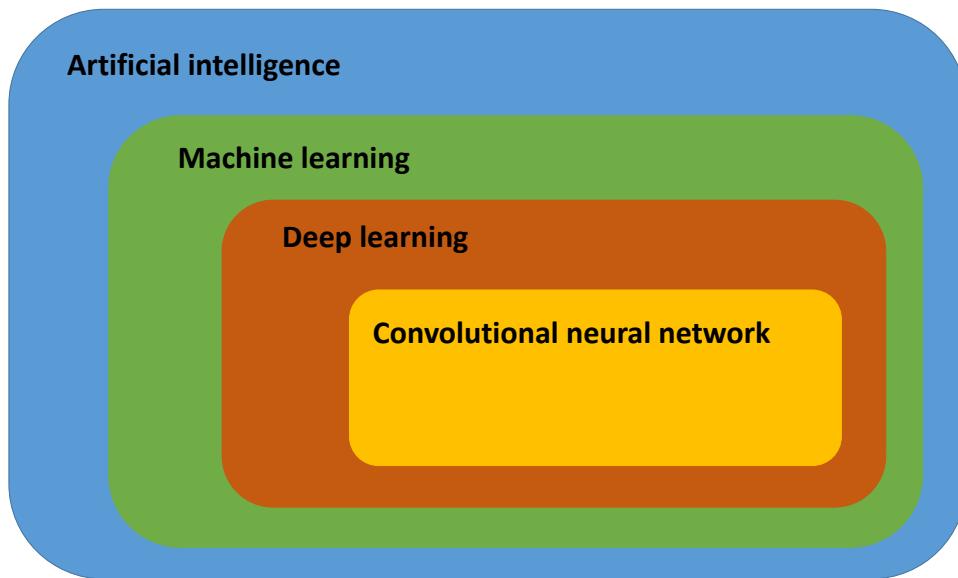


Figure 1: Conceptual framework of artificial intelligence with its subfields machine learning and deep learning.

are most related to the outcomes reduces the complexity of the model and increases training speed. Moreover, noisy and redundant features are eliminated which increases the performance of the model. In contrary to ML, neural networks can automatically select features. Therefore, DL can be trained directly on unstructured data like text, sound, video, and images. DL is a computationally expensive subfield of ML and requires large datasets to avoid generalization errors⁽⁹⁾. The number of neurons, number of layers, and connections between neurons determine the complexity and architecture of a DL algorithm⁽¹⁰⁾. The convolutional neural network (CNN) is a class of DL (Fig. 1) that is widely used for imaging applications. Trained CNNs have the ability to detect and classify distinctive features (e.g., edges of anatomical structures) on images, for example, to classify views of echocardiograms^(3, 11).

ML models are reported with a variety of metrics, which are selected for the ML application. Examples of metrics are the F1 score, accuracy, sensitivity, dice similarity coefficient (DSC), area under the receiver operating characteristic curve (AUC), and concordance statistic (C). These metrics are explained in detail elsewhere⁽¹²⁾.

Automated Interpretation of ICA

Search and Selection Strategy

A literature search was performed in the following databases: PubMed, Web of Science, Embase, and Google Scholar. The databases were searched in the publication period July 30, 2011 until July 30, 2021 with the following combined terms: (1) Coronary angiography AND (2) Artificial intelligence NOT (3) computed tomography. The exact search strategy is shown in the Appendix. Relevant studies were selected using machine learning–driven selection software called ASReview, which is further explained in the Appendix⁽¹³⁾. Relevant peer-reviewed articles were included if artificial intelligence models were developed on coronary angiography imaging data. Articles that solely focused on automated segmentation without other AI applications, reviews, and letters to editor were excluded. Records classified as non-relevant and reference lists were examined to find additional relevant studies. The search strategy resulted in 1335 studies. After deduplications and screening on title and abstract, 12 studies were included. A flow chart of study inclusion is shown in Supplementary Figure 1. The included studies reported on ML models for the following (diagnostic) applications: automated frame selection, segmentation, lesion assessment, and functional assessment of coronary flow. These applications will be summarized after a short introduction into ICA interpretation in daily clinical practice and its current limitations.

ICA Interpretation in Daily Clinical Practice

The interpretation of ICA is highly standardized and consists of the assessment of multiple components including coronary flow (Thrombolysis in Myocardial Infarction [TIMI] flow), lesion severity (percentage of stenosis and length) and other characteristics such as the presence of thrombi and calcifications. Despite standardized interpretation, ICA has well-known limitations. Coronary arteries are three-dimensional (3D) structures that are captured in two-dimensional (2D) images, which may result in overlap, foreshortening, and difficulty in assessing true (3D) stenosis grade. ICA image quality is further affected by low-dose radiation, commonly used in these procedures, heart motions, and X-ray absorbing tissues (e.g., ribs and vertebrae), which leads to low signal-to-noise ratio, low-contrast regions, and blur^(14–18). These limitations make ICA prone to subjective interpretation, which may have important diagnostic and therapeutic repercussions^(19, 20).

Frame Selection

ICA analysis is preferably performed during the end-diastolic phase of the cardiac cycle to minimize coronary artery motion and herewith prevent artifacts. Selection of contrast-filled frames in end-diastolic phase is a manual and time-consuming task, which lends itself for automation. Researchers demonstrated that a CNN could be trained on 56,655 coronary angiograms from 6820 patients to detect the end-diastolic phase⁽²¹⁾. Electrocardiography signals were used as ground truth. The CNN yielded good performance with an F1 score of 0.995. Instead of select-

ing one frame, other investigators trained a CNN on 90 ICA sequences in which three consecutive, contrast-filled frames were selected⁽²²⁾. The rationale of training with three consecutive frames was to reduce the number of false-positive observations of significant stenosis. These nonexistent stenoses are often visible on a single frame and caused by heterogeneity of contrast filling, curved vessels, coronary motion, or background noise. With an accuracy of 0.87 to select contrast-filled frames, the network performed better than conventional segmentation-based methods.

Segmentation

Selected frames can be segmented, which is a process to classify pixels as coronary arteries or irrelevant structures. Training an algorithm to identify relevant structures is crucial for detecting, localizing and classifying coronary lesions. To date, most studies on automated ICA image analysis have trained DL algorithms to automatically segment coronary arteries in coronary angiography^(15, 16, 23–28). Segmented coronary arteries can be partitioned into smaller structures based on, for example, location or anatomy. Recently, Du et al. trained a neural network (cGAN⁽²⁹⁾) on 12,323 angiograms collected from 2834 patients to label coronary arteries into 20 segments [⁽³⁰⁾]. Although not specified, this 20-segment model looks similar to the segment model of the globally accepted SYNTAX (Synergy Between Percutaneous Coronary Intervention With Taxus and Cardiac Surgery) score⁽³⁰⁾. The SYNTAX score is an objective tool to grade complexity of CAD and guides decision-making between PCI and coronary artery bypass graft (CABG). The recognition model was tested using an additional 1050 angiograms and showed a recognition accuracy of 98% and sensitivity of 85%. Both training and test data were collected from a single medical center. Aforementioned studies on automated frame selection and segmentation are shown in Table 1.

Despite limitations of ICA images on image quality, aforementioned studies show that it is feasible to train an AI algorithm to select frames of interest and automatically segment coronary arteries in a proper fashion.

Table 1: Summary of AI applications in intracoronary imaging

First author	Year	AI application	Data	Classifier	Metric value	Annotated by
Du T [30]	2021	Segmentation	20,612 ICA images of 10,073 patients	cGAN	ACC=98%, SE=85%	QA
Zhao C [42]	2021	Segmentation	314 ICA images of 99 patients	CNN	DSC=0.89	EC
Ciusdel C [21]	2020	End-diastolic frame detection	56,655 ICA sequences of 6820 patients	CNN	F1=99.5%	ECG
Wu W [22]	2020	Segmentation for frame selection	148 ICA sequences of 63 patients	CNN	SMT: visually, FS: ACC=0.87	EC
Moon JH [48]	2021	Lesion detection, localization, and classification	452 ICA images	CNN	AUC=0.96	QA and EC
Danilov VV [50]	2021	Lesion detection and localization	8325 ICA images of 100 patients	CNN	F1=0.96	EC

Continued on next page

First author	Year	AI application	Data	Classifier	Metric value	Annotated by
Du T [30]	2021	Lesion detection, localization, and classification	20,612 ICA images of 10,073 patients	CNN	F1=0.80–0.85	QA
Zhao C [42]	2021	Lesion detection, localization, and classification	314 ICA images of 99 patients	CNN	TPR=0.68, PPV=0.70	EC
Pang K [47]	2021	Lesion detection and localization	166 ICA sequences	CNN	F1=0.88	QA
Chen S [44]	2020	Lesion detection and classification	21,631 ICA sequences of 14,509 patients	CNN	F1=0.91–0.97	NS
Wu W [22]	2020	Lesion detection	148 ICA sequences of 63 patients	CNN	F1=0.83	EC
Yabushita H [46]	2020	Lesion detection	1838 ICA sequences 199 patients	CNN	C=0.61	EC
Ovalle-Magallanes E [49]	2020	Lesion detection	250 ICA images	CNN	F1=0.95	NS
Liu X [43]	2019	Lesion detection, localization, and classification	2059 ICA images	CNN	F1=0.89, AUC=0.98	EC
Roguin A [60]	2021	Fractional flow reserve estimation	31 patients	NS	SE=88%, SP=93%	EC
Cho H [59]	2019	Fractional flow reserve estimation	1717 patients	XGBoost	AUC=0.87	NS

ACC,accuracy; AUC,area under curve; cGAN,conditional generative adversarial network; CNN,convolutional neural network; DSC,dice similarity coefficient; EC, experienced cardiologist; F1, F1 score; ICA,invasive coronary angiography; NS,not specified; PPV,positive predictive value; SE,sensitivity; SP,specificity; SMT,segmentation; TPR,true-positive rate; QA,qualified analyst.

Lesion Detection, Localization, and Classifications

Several efforts have been made to improve ICA interpretation. Quantitative coronary angiography (QCA) software is already available for over three decades and can provide objective and quantitative assessment of anatomical lesion severity. However, QCA requires manual, time-consuming input and has therefore not been widely implemented into clinical practice^(31, 32). In recent years, software has been developed to reduce noise and improve detection of stenosis in coronary arteries^(14, 33–36). However, these methods are often computationally expensive^(22, 37–39), semi-automatic, and have long processing times^(22, 39–41).

The inter- and intra-observer variability of visual assessment of lesions by clinicians could be minimized if lesion detection, localization, and classification are automated. Du and colleagues trained a CNN on 6239 lesions to improve lesion detection and to categorize lesions into stenotic lesions, total occlusions, calcific lesions, and the presence of thrombus or dissection. Internal validation of CNN performance on 1000 ICA images demonstrated F1 scores between 0.80 and 0.85. Other studies performed classification on the degree of stenosis (mild, moderate, severe), or elements of SYNTAX, such as the presence and type (blunt/tapered stump) of total occlusion with moderate to good results (Table 1)^(42–44).

Large amounts of labeled data are needed to train an algorithm that generalizes well to unseen data⁽⁴⁵⁾. In a study by Yabushita et al. training on 199 ICA images resulted in modest performance

(C=0.61) to detect the presence of clinically significant coronary stenoses⁽⁴⁶⁾. In the setting of lower volume datasets, diagnostic accuracy of ML models could be enhanced by several strategies. As an example, training of CNNs on sequences of frames improved the rate of false-positive stenoses. Researchers demonstrated that by incorporating temporal information, F1 scores increased by 30–40%^(22, 47). Transfer learning and data augmentation are other strategies that can increase performance. In transfer learning, an AI model, already being trained for another task, will be further made ready for other purposes. Data augmentation is a technique to increase the amount of data without collecting new data. For example, an AI model pre-trained on a large image database (ImageNet) was further trained on 45,125 frames to localize stenoses with a cut-off of 50% in the right coronary artery⁽⁴⁸⁾. The ICA frames were derived from 452 ICA frames by cropping and pixel intensity value adjustments. Validation on an external dataset yielded an excellent AUC of 0.96. Other researchers also demonstrated the power of pre-training by employing an automated lesion detection CNN by means of training on only 125 images. Despite the limited amount of training data, F1 score was as high as 0.95⁽⁴⁹⁾.

Real-time detection of coronary stenoses can facilitate operators to identify lesions that might have otherwise been unnoticed. However, the processing time of such an AI model is an important constraint which is often not reported in studies. As a fact, a better ML model accuracy often means a higher complexity of its architecture and processing time⁽⁵⁰⁾. Real-time application of AI should not result in time delays, which may affect the outcome of patients⁽⁵¹⁾. Therefore, there should be a trade-off between accuracy and speed in deployed models⁽⁵⁰⁾.

In summary, there has been a great deal of progression in automated detection and classification of CAD in the last decade. These developments are attributable to gains in computing power, advances in ML algorithms, and availability of large ICA datasets⁽⁵²⁾. Automated detection and classification of CAD may provide physicians objective and reproducible information and may prevent significant lesions to be missed^(22, 30, 50). All studies on automated lesion detection, localization, and classification are shown in Table 1.

Functional Assessment of Coronary Flow

Functional sufficiency of coronary flow and plaque characterization are fundamental elements that guide treatment decisions but cumbersome features to assess on ICA^(20, 53). There is a discrepancy between visual assessment and intracoronary pressure measurement for assessment of functional sufficiency of coronary flow^(54, 55). Therefore, intracoronary pressure measurements are performed to assess whether a stenosis is functionally significant and herewith causes myocardial ischemia⁽⁵⁶⁾. Fractional flow reserve (FFR) is the most used metric and records the mean distal coronary pressure divided by the mean proximal pressure during maximal hyperemia. Although evidence shows that FFR-based decision-making for revascularization leads to improved cardiovascular outcomes^(57, 58), the FFR technique has its limitations. Major limitations of this technique are its invasive nature and necessity of use of costly pressure wires. In addition, prolonged procedural time and operator's preference for visual assessment limit the implementation of routine intracoronary measurements during ICA^(54, 55, 59, 60). To overcome these

limitations, quantitative flow ratio (QFR) applications have been developed to add functional assessment to anatomic imaging analysis. QFR is a non-invasive method to calculate functional sufficiency based on 3D-angiographic reconstruction and computational fluid dynamics^(61, 62). To date, QFR analysis is not readily available for daily clinical practice at the catheterization laboratory (cath lab) and requires computationally expensive post-processing.

AI-based FFR estimation is likely to have less processing time compared to QFR estimation based on computational fluid dynamics, as demonstrated by studies on AI-based FFR estimation on coronary CT^(63, 64). Recently, a ML model (XGBoost⁽⁶⁵⁾) was developed on data of 1501 patients to classify intermediate lesions as having a $\text{FFR} \leq 0.8$ or $\text{FFR} > 0.8$ ⁽⁵⁹⁾. Feature selection resulted in a set of 12 features including body surface area, sex, and 10 features extracted from ICA images (lengths and diameters of lumen and stenosis). Evaluation of this classification model on an external dataset yielded an AUC of 0.87. More recently, a feasibility study was conducted to compare novel AI-based FFR software to invasive FFR measurements⁽⁶⁰⁾. This software, called AutocathFFR, was able to detect coronary lesions and predict their FFR value without coronary artery annotation or view selection. The diagnostic value of AutocathFFR to classify a lesion as functional significant was evaluated in 31 patients, with the left anterior descending artery as the most frequent target (25 of 31 patients). The sensitivity, specificity, positive predictive value, and negative predictive value were 0.88, 0.93, 0.94, and 0.87, respectively. These results are similar to the performance of QFR and demonstrate the feasibility for automated FFR estimation. All studies on AI-based functional assessment of coronary flow are shown in Table 1.

Although automated assessment of FFR directly from ICA images has potential to speed up procedures, studies investigating real-time QFR-based PCI versus standard of care (i.e., FFR-guided PCI) are still ongoing. Successful introduction of QFR-based coronary treatment might eventually reduce over- and under-treatment. Furthermore, the need to perform intracoronary hemodynamic measurements will diminish, which might result in lower incidence of complications and lower healthcare costs.

Limitations and Challenges in Development of Automated ICA Analysis

AI has the potential to increase diagnostic performance and support clinicians in therapeutic decision-making by automatically assessing the extent and functional significance of CAD in the cath lab. However, multiple barriers have to be overcome before these models can be implemented into clinical practice.

AI Bias

A key challenge in development of smart technology is to work toward generalizable AI applications, which are externally validated and trained on large and variable patient populations from multiple centers⁽⁶⁶⁾. However, because most studies are proof-of-concept studies, external validation is often not performed⁽⁶⁷⁾. Results of only three out of the 12 studies mentioned in Table 1 have been externally validated^(46, 48, 59). Therefore, datasets should be shared between research

centers or made open-access according to FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles⁽⁶⁸⁾. To avoid the risk of algorithmic bias, subgroup analysis on populations (e.g., age, ethnicity, sex, and medical center) should be performed. These analyses will show whether population subgroups were underrepresented in the training data and whether more data for training should be collected⁽⁶⁶⁾.

AI Interpretability

Diagnosis and therapeutic decision making has a tremendous impact on clinical outcomes in the cath lab. There is a need for AI applications in which algorithmic decisions are clearly explained (explainable AI), so that eventual inaccurate analysis can be back traced. However, algorithmic decisions are often difficult to understand due to its “black box” nature. Currently, ongoing research on explainable AI is likely to enhance trust among users and facilitate adoption of AI applications^(69–71).

AI in ICA versus Other Cardiac Modalities

To our best knowledge, no studies have been published on AI applications developed for treatment guidance, risk stratification, or prognosis based on ICA imaging. This is in contrast to the cardiac imaging modalities echocardiography, coronary computed tomography angiography (CCTA) and magnetic resonance imaging (MRI), in which the first AI applications to predict prognosis have emerged^(5, 72–76). This development delay could be explained by several factors. In ICA, registration of 3D structures in 2D images results in overlapping coronary arteries, which hinder AI models to find coronary artery specific features. This overlap in coronary arteries with anatomical variation and heterogeneity among operators regarding X-ray beam projections results in the need for large ICA datasets in order to develop a well-performing AI model in the setting of ICA. Pre-processing (selection, segmentation, classification) of these datasets is a time-consuming and tedious process. Another possible explanation is that ICA has a different role in the diagnostic and treatment trajectory of patients with (suspected) ischemic heart disease compared to other, noninvasive imaging modalities. Its invasive nature makes ICA a second-line diagnostic test, only applied in patients with high a priori probability of CAD. This might favor development of AI models for other cardiac imaging modalities earlier in the diagnostic trajectory compared to ICA.

Future Perspective of AI in the Cath Lab

Growing Healthcare Utilization

ICA has numerous important challenges to overcome in the next decades. Growing burden of cardiovascular disease is likely to increase the number of interventions being performed. As a consequence, workload and healthcare costs will further increase. Without smart solutions, personnel exhaustion and delayed or cancelled interventions will jeopardize quality of care⁽⁷⁷⁾.

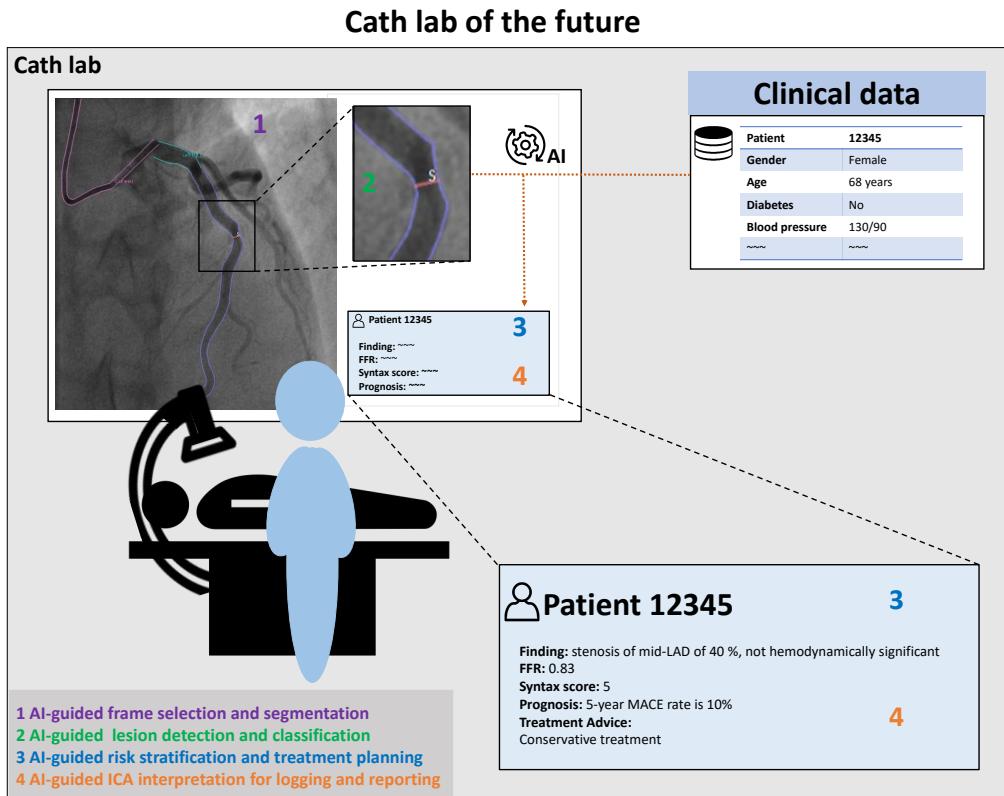


Figure 2: Example of future, automated invasive coronary angiography analysis: artificial intelligence (AI) for automated quantitative coronary angiography (QCA) with FFR estimation, (syntax-based) clinical risk scoring and reporting.

AI as a smart technology has the potential to alleviate pressure on healthcare services in general and to improve cath lab diagnoses, treatment, and logistics in particular.

AI-Guided ICA Interpretation for Logging and Reporting

The amount of administrative work of employees of the cath lab is increasing swiftly and ensures that time available for efficient patient care is minimized⁽⁷⁸⁾. Automated logging and reporting of procedures by automated, AI-based ICA interpretation can reduce this administrative burden (see Fig. 2). Some examples of mundane tasks that may be reported automatically are the location and significance of the lesion and whether implants (e.g., stents) have been placed.

AI-Guided Treatment Planning

AI models that allow accurate and fast evaluation of coronary anatomy and noninvasive functional sufficiency will offer an opportunity to further develop AI technology that will be able to guide real-time PCI procedures. Peri-procedural analysis of ICA images, including automated functional assessment, could optimize PCI outcomes by providing a lesion-specific recommendation on a revascularization strategy, eventually with advice on stent size, length, location, and preferred strategy (Fig. 2). After stenting, automated measurements on the proportion of stent under expansion and hemodynamic function may inform the operator and patient about the expected short- and long-term outcome^(79, 80).

AI-Guided Risk Stratification and Prognosis

The SYNTAX score, and subsequently the SYNTAX II score (which adds clinical characteristics to the anatomical assessment of the coronary tree), are examples of available stratification tools to guide clinical decision-making in complex CAD. However, these scores are time-consuming (5–10 min) to calculate and therefore underutilized in daily clinical practice, especially during ICA. Improved SYNTAX-like scores, integrating automated AI-based ICA imaging analysis and key clinical characteristics (extracted from electronic patient dossiers by intelligent and complex AI applications), might improve risk stratification of the individual patient and herewith enhance patient-tailored treatment, and ultimately prognosis (Fig. 2).

Other AI-Guided Applications

Other AI applications in ICA, beyond the scope of this review, may reduce radiation exposure by focusing on image acquisition and reconstruction⁽⁸¹⁾. In addition, AI in intracoronary imaging (e.g., intravascular ultrasound (IVUS), optical coherence tomography (OCT), near-infrared spectroscopy (NIRS)) may lead to improved identification of truly vulnerable coronary plaques and may further elucidate the genesis of in-stent restenosis⁽⁸²⁾. Ultimately, AI-based integration of upstream diagnostic modalities (i.e., CCTA), ICA, and intracoronary imaging may lead to optimal outcomes after PCI.

Conclusion

The cath lab is on the verge of a new era in which AI-based state-of-the-art models are being developed for diagnostic and treatment guidance, optimized risk stratification, and automated cath lab logistics. We are still in an early stage of development, as most models are constructed on single-center datasets and external validation is often lacking. Large multicenter datasets are necessary to develop more generalizable models and cath lab field-labs, mirrored to real-life cath labs, are indispensable to readily test them.

References

- [1] Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nat Rev Cardiol.* Nature Publishing Group; 2021;1–10.
- [2] Leiner T, Rueckert D, Suinesiaputra A, Baeßler B, Nezafat R, İlsgüm I, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *J Cardiovasc Magn Reson.* 2019;21:61.
- [3] Schuurings MJ, İlsgüm I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med.* 2021;8:648877.
- [4] Opincariu D, Benedek T, Chițu M, Raț N, Benedek I. From CT to artificial intelligence for complex assessment of plaque-associated risk. *Int J Cardiovasc Imaging.* 2020;36:2403–27.
- [5] Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* 2017;38:500–7.
- [6] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J.* 2020;41:407–77.
- [7] Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–30.
- [8] Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, et al. Clinical phenogroups in heart failure with preserved ejection fraction: detailed phenotypes, prognosis, and response to spironolactone. *JACC Heart Fail.* 2020;8:172–84.
- [9] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol.* 2019;19:64.
- [10] Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinformatics.* 2018;16:17–32.
- [11] Howard JP, Francis DP. Machine learning with convolutional neural networks for clinical cardiologists. *Heart.* 2022;108(12):973–981.
- [12] Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *Am J Roentgenol. American Roentgen Ray Society.* 2019;212:38–43.
- [13] van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdem F, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell.* 2021;3:125–33.
- [14] Fazlali HR, Karimi N, Soroushmehr SMR, Sinha S, Samavi S, Nallamothu B, et al. Vessel region detection in coronary X-ray angiograms. *2015 IEEE Int Conf Image Process (ICIP).* 2015;2015:1493–7.
- [15] Cervantes-Sánchez F, Cruz-Aceves I, Hernandez-Aguirre A, Hernandez-Gonzalez MA, Solorio-Meza SE. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Appl Sci.* 2019;9:5507.
- [16] Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Ward K, Jafari MH, et al. Vessel extraction in X-ray angiograms using deep learning. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2016:643–6.
- [17] Cruz-Aceves I, Cervantes-Sánchez F, Avila-Garcia MS. A Novel Multiscale Gaussian-Matched Filter Using Neural Networks for the Segmentation of X-Ray Coronary Angiograms. *J Healthc Eng.* 2018.
- [18] Kobayashi T, Hirshfeld JW. Radiation exposure in cardiac catheterization. *Circulation: Cardiovascular Interventions. American Heart Association.* 2017;10:e005689.
- [19] Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver variability in coronary angiography. *Circulation.* 1976;53:627–32.
- [20] Lee CH, Hur S-H. Optimization of percutaneous coronary intervention using optical coherence tomography. *Korean Circ J.* 2019;49:771–93.
- [21] Ciudel C, Turcea A, Puiu A, Itu L, Calmac L, Weiss E, et al. Deep neural networks for ECG-free cardiac phase and end-diastolic frame detection on coronary angiographies. *Comput Med Imaging Graph.* 2020;84:101749.
- [22] Wu W, Zhang J, Xie H, Zhao Y, Zhang S, Gu L. Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. *Comput Biol Med.* 2020;118:103657.
- [23] Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep. Nature Publishing Group.* 2019;9:1–11.
- [24] Shin SY, Lee S, Yun ID, Lee KM. Deep vessel segmentation by learning graphical connectivity. *Med Image Anal.* 2019;58:101556.
- [25] Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, et al. Multichannel fully convolutional network for coronary artery segmentation in X-ray angiograms. *IEEE Access.* 2018;6:44635–43.
- [26] Zhu X, Cheng Z, Wang S, Chen X, Lu G. Coronary angiography image segmentation based on PSPNet. *Comput Methods Programs Biomed.* 2020.
- [27] Jo K, Kweon J, Kim Y-H, Choi J. Segmentation of the main vessel of the left anterior descending artery using selective feature mapping in coronary angiography. *IEEE Access.* 2019;7:919–30.
- [28] Jun TJ, Kweon J, Kim Y-H, Kim D. T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography. *Neural Netw.* 2020;128:216–33.
- [29] Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv:14111784.* 2014.
- [30] Du T, Xie L, Zhang H, Liu X, Wang X, Chen D, et al. Automatic and multimodal analysis for coronary angiography: training and validation of a deep learning architecture. *EuroIntervention.* 2020;17(1):32–40.
- [31] Zhang H, Mu L, Hu S, Nallamothu BK, Lansky AJ, Xu B, et al. Comparison of physician visual assessment with quantitative coronary angiography in assessment of stenosis severity in China. *JAMA Intern Med.* 2018;178:239–47.
- [32] Nallamothu BK, Spertus JA, Lansky AJ, Cohen DJ, Jones PG, Kureshi F, et al. Comparison of clinical interpretation with visual assess-

- ment and quantitative coronary angiography in patients undergoing percutaneous coronary intervention in contemporary practice. *Circ Am Heart Assoc.* 2013;127:1793–800.
- [33] Nakamura S, Kobayashi T, Funatsu A, Okada T, Mauti M, Waizumi Y, et al. Patient radiation dose reduction using an X-ray imaging noise reduction technology for cardiac angiography and intervention. *Heart Vessels.* 2016;31:655–63.
- [34] Ten Cate T, van Wely M, Gehlmann H, Mauti M, Camaro C, Reifart N, et al. Novel X-ray image noise reduction technology reduces patient radiation dose while maintaining image quality in coronary angiography. *Neth Heart J.* 2015;23:525–30.
- [35] Nirmala Devi S, Kumaravel N. Comparison of active contour models for image segmentation in X-ray coronary angiogram images. *J Med Eng Technol.* 2008;32:408–18.
- [36] Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms — review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed.* 2018;158:71–91.
- [37] Cruz-Aceves I, Oloumi F, Rangayyan RM, Aviña-Cervantes JG, Hernandez-Aguirre A. Automatic segmentation of coronary arteries using Gabor filters and thresholding based on multiobjective optimization. *Biomed Signal Process Control.* 2016;25:76–85.
- [38] Fazlali HR, Karimi N, Soroushmehr SMR, Shirani S, Nallamothu BK, Ward KR, et al. Vessel segmentation and catheter detection in X-ray angiograms using superpixels. *Med Biol Eng Comput.* 2018;56:1515–30.
- [39] Iyer K, Najarian CP, Fattah AA, Arthurs CJ, Soroushmehr SMR, Subban V, et al. AngioNet: a convolutional neural network for vessel segmentation in X-ray angiography. *Sci Rep.* 2021;11(1):18066.
- [40] Brieva J, Galvez M, Toumoulin C. Coronary extraction and stenosis quantification in X-ray angiographic imaging. *Conf Proc IEEE Eng Med Biol Soc.* 2004;2004:1714–7.
- [41] Fatemi MJR. Detection of narrowed coronary arteries in X-ray angiographic images using contour processing of segmented heart vessels based on Hessian vesseness filter and wavelet based image fusion. *Int J Comput Appl.* 2011;36:27–33.
- [42] Zhao C, Vij A, Malhotra S, Tang J, Tang H, Pienta D, et al. Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Comput Biol Med.* 2021;136:104667.
- [43] Liu X, Du T, Zhang H, Sun C. Detection and classification of chronic total occlusion lesions using deep learning. *Annu Int Conf IEEE Eng Med Biol Soc.* 2019;2019:828–31.
- [44] Chen S, Tang Y, Shi X, Zhang H, Xie L, Xu B. Convolution pyramid network: a classification network on coronary artery angiogram images. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020;2020:1186–9.
- [45] Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet Elsevier.* 2018;392:2388–96.
- [46] Yabushita H, Goto S, Nakamura S, Oka H, Nakayama M, Goto S. Development of novel artificial intelligence to detect the presence of clinically meaningful coronary atherosclerotic stenosis in major branch from coronary angiography video. *J Atheroscler Thromb.* 2021;28(8):835–843.
- [47] Pang K, Ai D, Fang H, Fan J, Song H, Yang J. Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography. *Comput Med Imaging Graph.* 2021;89:101900.
- [48] Moon JH, Lee DY, Cha WC, Chung MJ, Lee K-S, Cho BH, et al. Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Comput Methods Programs Biomed.* 2021;198:105819.
- [49] Ovalle-Magallanes E, Avina-Cervantes JG, Cruz-Aceves I, Ruiz-Pinales J. Transfer learning for stenosis detection in X-ray coronary angiography. *Mathematics.* 2020;8:1510.
- [50] Danilov VV, Klyshnikov KY, Gerget OM, Kutikhin AG, Ganyukov VI, Frangi AF, et al. Real-time coronary artery stenosis detection based on modern neural networks. *Sci Rep.* 2021;11:7582.
- [51] De Luca G, Suryapranata H, Ottervanger JP, Antman EM. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. *Circulation.* 2004;109:1223–5.
- [52] Ramkumar PN, Kunze KN, Haeblerle HS, Karnuta JM, Luu BC, Nwachukwu BU, et al. Clinical and research medical applications of artificial intelligence. *Arthroscopy.* 2021;37:1694–7.
- [53] Mintz GS, Popma JJ, Pichard AD, Kent KM, Satler LF, Chien Chuang Y, et al. Limitations of angiography in the assessment of plaque distribution in coronary artery disease. *Circulation American Heart Association.* 1996;93:924–31.
- [54] Rigattieri S, Biondi Zocca G, Sciahbasi A, Di Russo C, Cera M, Patrizi R, et al. Meta-Analysis of head-to-head comparison of intracoronary versus intravenous adenosine for the assessment of fractional flow reserve. *Am J Cardiol.* 2017;120:563–8.
- [55] Park S-J, Kang S-J, Ahn J-M, Shin EB, Kim Y-T, Yun S-C, et al. Visual-functional mismatch between coronary angiography and fractional flow reserve. *JACC Cardiovasc Interv.* 2012;5:1029–36.
- [56] Neumann F-J, Sousa-Uva M, Ahlsson A, Alfonso F, Banning AP, Benedetto U, et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J.* 2019;40:87–165.
- [57] Tonino PAL, De Bruyne B, Pijls NHJ, Siebert U, Ikeno F, van't Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med.* 2009;360:213–24.
- [58] Pijls NHJ, Fearon WF, Tonino PAL, Siebert U, Ikeno F, Bornschein B, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention in patients with multivessel coronary artery disease: 2-year follow-up of the FAME (Fractional Flow Reserve Versus Angiography for Multivessel Evaluation) study. *J Am Coll Cardiol.* 2010;56:177–84.
- [59] Cho H, Lee J, Kang S, Kim W, Choi S, Ko J, et al. Angiography-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesions. *J Am Heart Assoc [Internet].* 2019;8(4):e011685.
- [60] Roguin A, Abu Dogosh A, Feld Y, Konigstein M, Lerman A, Koifman E. Early feasibility of automated artificial intelligence angiography based fractional flow reserve estimation. *Am J Cardiol.* 2021;139:8–14.

- [61] Emori H, Kubo T, Kameyama T, Ino Y, Matsuo Y, Kitabata H, et al. Quantitative flow ratio and instantaneous wave-free ratio for the assessment of the functional severity of intermediate coronary artery stenosis. *Coron Artery Dis.* 2018;29:611–7.
- [62] Morris PD, Silva Soto DA, Feher JFA, Rafroiu D, Lungu A, Varma S, et al. Fast virtual fractional flow reserve based upon steady-state computational fluid dynamics analysis: results from the VIRTU-Fast Study. *JACC: Basic Transl Sci.* 2017;2:434–46.
- [63] Coenen A, Kim Y-H, Kruk M, Tesche C, De Geer J, Kurata A, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE Consortium. *Circ Cardiovasc Imaging.* 2018;11(6):e007217
- [64] Itu L, Rapaka S, Passerini T, Georgescu B, Schwemmer C, Schoebinger M, et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *J Appl Physiol.* 2016;121:42–52.
- [65] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;785–94.
- [66] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.
- [67] Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol.* 2019;20:405–10.
- [68] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
- [69] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci U S A.* 2019;116:22071–80.
- [70] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis.* 2020;128:336–59.
- [71] Ras G, Xie N, van Gerven M, Doran D. Explainable Deep Learning: A Field Guide for the Uninitiated. arXiv:200414545, 2021.
- [72] van Rosendael AR, Maliakal G, Kolli KK, Beecy A, Al'Aref SJ, Dwivedi A, et al. Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry. *J Cardiovasc Comput Tomogr.* 2018;12:204–9.
- [73] de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Isgrum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans Med Imaging.* 2019;38:2127–38.
- [74] Kwon J-M, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography.* 2019;36:213–8.
- [75] Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, et al. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. *JACC Cardiovasc Imaging.* 2019;12:681–9.
- [76] Bello GA, Dawes TJW, Duan J, Bif C, de Marvao A, Howard LSGE, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nat Mach Intell.* 2019;1:95–104.
- [77] Wen J, Cheng Y, Hu X, Yuan P, Hao T, Shi Y. Workload, burnout, and medical mistakes among physicians in China: a cross-sectional study. *Biosci Trends.* 2016;10:27–33.
- [78] Woolhandler S, Himmelstein DU. Administrative work consumes one-sixth of U.S. physicians' working hours and lowers their career satisfaction. *Int J Health Serv.* 2014;44:635–42.
- [79] Kang S-J, Mintz GS, Park D-W, Lee S-W, Kim Y-H, Whan Lee C, et al. Mechanisms of in-stent restenosis after drug-eluting stent implantation. *Circulation: Cardiovascular Interventions.* Am Heart Assoc. 2011;4:9–14.
- [80] Koo B-K, Samady H. Strap in for the artificial intelligence revolution in interventional cardiology. *JACC: Cardiovascular Interventions.* 2019;12:1325–7.
- [81] Bang JY, Hough M, Hawes RH, Varadarajulu S. Use of Artificial intelligence to reduce radiation exposure at fluoroscopy-guided endoscopic procedures. *Am J Gastroenterol.* 2020;115:555–61.
- [82] Fedewa R, Puri R, Fleischman E, Lee J, Prabhu D, Wilson DL, et al. Artificial intelligence in intracoronary imaging. *Curr Cardiol Rep.* 2020;22:46.

Supplementary material

Search Strategy

Pubmed

("Coronary Angiography"[Mesh] OR "coronary angiogra*" [tiab] OR "Coronary Vessels"[Mesh] OR "Coronary Stenosis"[Mesh] OR "coronary stenosis" [tiab]) AND ("Artificial Intelligence"[Mesh] OR "Artificial Intelligence" [tiab] OR "AI" [tiab] OR "machine learning" [tiab] OR "deep learning" [tiab] OR "neural network" [tiab] OR "CNN" [tiab] OR "automatic image analysis" [tiab] OR "computer-aided diagnosis" [tiab]) NOT ("CT" [ti] OR "tomogr*" [ti])

Web of Science Search

TS=("Coronary Angiography" OR "coronary angiogra*" OR "Coronary Vessels" OR "coronary stenosis") AND TI=("Artificial Intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "neural network" OR "CNN" OR "automatic image analysis" OR "computer-aided diagnosis") NOT TI=("CT" OR "tomogr*")

Embase

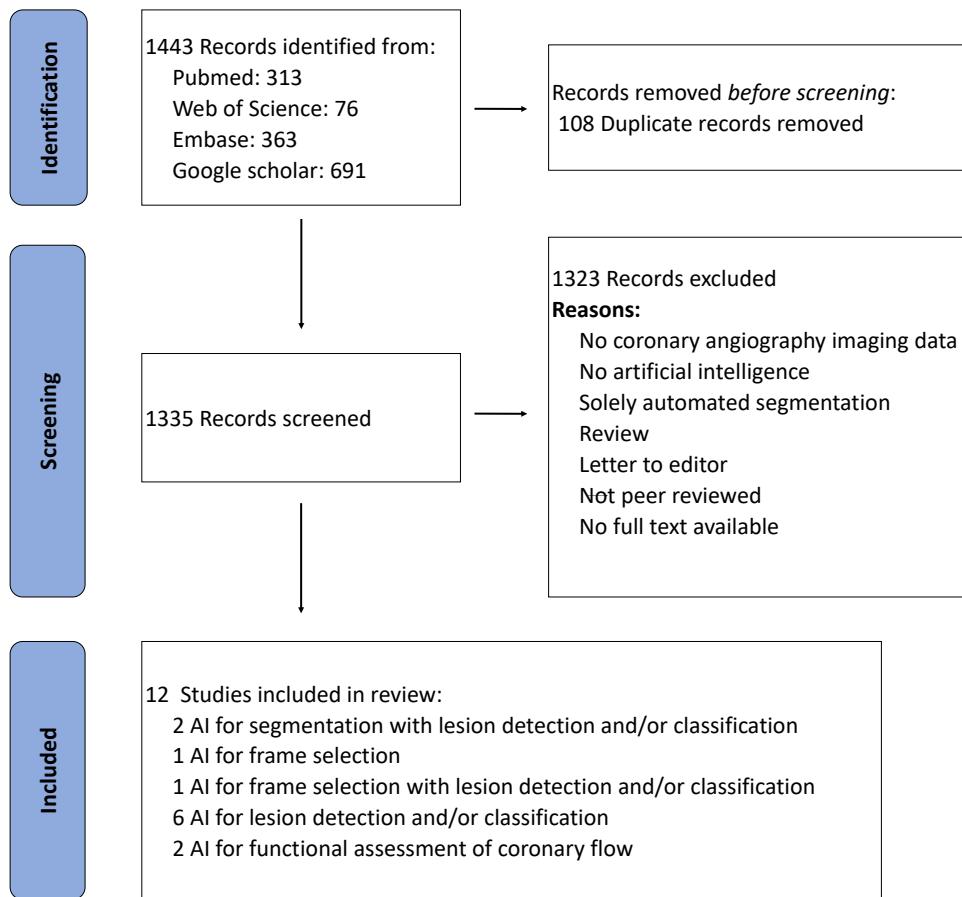
((coronary angiography OR coronary vessels OR coronary sten*).ti,ab OR exp coronary angiography/) and (exp artificial intelligence/ or artificial intelligence.ti,ab,kw. or exp deep learning/ or deep learning.ti,ab,kw. or exp machine learning/ or machine learning.ti,ab,kw. or exp artificial neural network/ or artificial intelligence.ti,ab,kw.) NOT (ct OR computed tomography OR computed tomographic angiography).ti

Google Scholar

"coronary angiography" AND ("artificial intelligence" | "deep learning") -"ct" -"tomography" The search was limited to studies written in English, human studies, and articles published in a peer-reviewed journal. The titles and abstracts of full text available articles were assessed on eligibility by a single researcher.

Study selection

ASReview selects relevant records by active learning, which is an interactive tool to train an algorithm with less data. Feature selection was performed with a natural language processing method called frequency-inverse document frequency. A Naïve Bayes classifier was iteratively trained to label records as relevant and non-relevant.



Supplementary Figure 1: Flowchart of study inclusion.

Deep learning-based segmentation of coronary arteries and stenosis detection in x-ray coronary angiography

Mitchel A. Molenaar

Elsa Hebbo

Jasper L. Selder

Nicoloz Shekiladze

Pratik B. Sandesara

William J. Nicholson

Folkert W. Asselbergs

Syed Ahmad

Daniel A. Gold

Shaimaa M. Sakr

Javier Oliván Bescós

Vincent Auvray

Martijn S. van Mourik

Alexander Haak

Yida Zhao

Jelle D. Nieuwendijk

Mark J. Schuuring

Berto J. Bouma

Steven A.J. Chamuleau

Niels J. Verouden

Abstract

Background

Deep learning applications may assist in automatically detecting coronary arteries on invasive coronary angiography (ICA).

Objectives

We aimed to train deep learning models for the segmentation of coronary arteries and the detection of significant stenosis on ICA, conduct external validation, and compare the performance with expert variabilities.

Methods

ICA studies from Amsterdam University Medical Centers (Center 1) and Emory University Hospital (Center 2) were retrospectively collected. Contours of the main coronary arteries and their $\geq 50\%$ stenoses were manually segmented using dedicated software. Deep learning-based models were created using data from Center 1, Center 2, and both centers. The performance of the models was assessed on unseen data and compared to expert variability.

Results

A total of 10,573 ICA images were used to train models: 9065 from Center 1 (2624 patients) and 1508 (456 patients) from Center 2. Validation was done on 186 Center 1 images and 123 Center 2 images. The segmentation model trained on datasets from both centers had the highest median Dice coefficient (0.84). The stenoses detection algorithm trained on both centers achieved a rate of 64%, similar to the expert agreement (65%). The model trained on the data with the most stenoses yielded the highest stenosis detection rate (66%). When matched in size and in number of stenosis, the models trained on both centers perform similarly.

Conclusion

The models achieved performance levels on par with experts in the segmentation of coronary arteries and detection of significant stenosis in the main arteries.

Introduction

Coronary artery disease (CAD) is the narrowing of the coronary arteries, which is the most common cardiovascular disease leading to more than 9 million deaths annually⁽¹⁾. In patients with suspected CAD, invasive coronary angiography (ICA) is the standard diagnostic evaluation performed in intermediate to high-risk patients^(2, 3). This diagnostic imaging procedure offers accurate characterization of the coronary anatomy and the severity of stenoses, providing physicians with essential information for making treatment decisions, either medical therapy or percutaneous or surgical revascularization⁽³⁾.

ICA has several limitations. The assessment of the two-dimensional ICA images is impeded by artery foreshortening, artery overlap, and low image quality⁽⁴⁾. In addition, the severity of stenosis is often assessed visually, even though numerous studies have shown that visual assessment of stenosis severity is limited by inaccuracy and inter- and intra-observer variability^(5, 7). Previous studies have reported that the severity of stenosis is overestimated by visual assessment compared to quantitative coronary angiography (QCA) by up to 21%⁽⁵⁾, which may affect the decision to opt for additional invasive intravascular imaging and/or physiologic assessment⁽³⁾, and revascularization⁽⁸⁾. Although QCA software can provide quantitative anatomical stenosis severity assessment, it is rarely used due to tedious user interactions including frame selection and calibration⁽⁹⁾. To transfer QCA from a mainly offline research tool to the catheterization room, user-friendly software is needed that offers real-time support during procedures with minimal manual input⁽¹⁰⁾.

Deep learning, a subtype of machine learning, is increasingly employed in medical imaging due to its ability to extract information from images that previously required human intelligence^(11, 12). Recent studies have demonstrated that deep learning-applications may assist physicians in the interpretation of ICA images. These applications mainly focused on automated segmentation (detection) of coronary arteries^(10, 13–20), which is a prerequisite for the automated stenosis assessment^(9, 21). However, most studies did not validate their segmentation models on data from a different institution, thus possibly introducing intrinsic selection biases of operators, x-ray systems, data storage, and annotators⁽¹⁰⁾. To assess the potential for generalization of these models to various clinical settings, this study aimed to train deep learning models for the segmentation of coronary arteries and the detection of significant stenosis on ICA, and to conduct external validation.

Methods

Study Population and Image Acquisition

Studies of patients who underwent ICA or percutaneous coronary intervention were retrospectively collected from the archiving systems of two tertiary centers (Center 1: Amsterdam UMC, The Netherlands; Center 2: Emory University School of Medicine, Atlanta, United States of America) between 2015 and 2017 (Center 1), and 2005 to 2021 (Center 2). Patients with a history of coronary artery bypass surgery were excluded. Operators acquired the ICA through Philips equipment (Philips Medical Systems, Best, The Netherlands).

A total of 2673 patients with 9065 images from center 1, and 456 ICA patients with 1508 images from center 2, were selected. The ICA studies were extracted from the picture archiving and communication system (PACS), in which the studies were stored in the 512 x 512 Digital Imaging and Communications in Medicine (DICOM) format, and anonymized. Each study encompassed multiple ICA cine runs, which were captured from different views by the operator to perform a comprehensive evaluation of the coronary arteries. This study complied with the principles in the Declaration of Helsinki, and received approval by the local human ethical review board. The study met the criteria for a waiver of the informed consent requirements.

Data Annotation

To train a deep learning-based segmentation model, a set of segmentations were created in both centers. For each ICA study, end-diastolic frames from multiple views were selected. A standardized method was employed to identify end-diastolic frames. Annotators examined multiple frames and selected the one where the coronary arteries were maximally dilated, defining it as the end-diastolic frame. The optimal frames to assess the extent and severity of the significant stenosis (stenosis degree $\geq 50\%$) were then selected. If the patient did not have any significant stenoses, the optimal frame to label the coronary anatomy was selected. In general, two runs were selected for the left coronary tree and one run was selected for the right coronary tree.

The left circumflex artery (LCX), left anterior descending artery (LAD), and right circumflex artery (RCA) were manually annotated using dedicated software according to the segment model of the globally accepted SYNTAX (Synergy Between Percutaneous Coronary Intervention With Taxus and Cardiac Surgery⁽²²⁾) score. All visible significant stenoses (quantifiable stenosis degree $\geq 50\%$), which were not impaired by low contrast or significant overlap, were annotated by drawing a line at the minimal lumen diameter. Reference lines were drawn in the coronary arteries proximal and distal from the stenosis.

In Center 1, the ICA images were selected and annotated by ten medical students who underwent comprehensive training by interventional cardiologists. The accompanied reports were verified for the presence of stenoses. The images were randomly reviewed by three experienced annotators (an interventional cardiologist, a cardiologist, and a technical physician), and corrections were made if necessary. In Center 2, the images were selected and annotated by two medical students who were trained by two interventional cardiologists. The work was supervised by

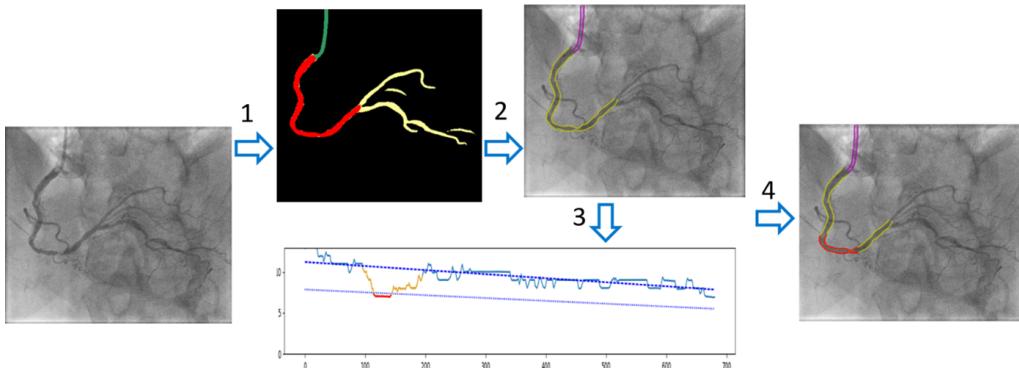


Figure 1: Model Input and Output. 1. An artificial intelligence-based segmentation of the main arteries is performed. 2. The artery borders are extracted and refined. 3. The corresponding diameter profile is analyzed. 4. Potentially detected stenosed segments are reported and automatically measured.

the interventional cardiologists.

A total of 171 images from 49 patients were randomly selected as a test set from Center 1, and an additional 109 images from 32 patients were selected from Center 2. These unseen data were used for calculating the final quantitative and qualitative assessments.

Segmentation and Stenosis Detection Model

A model was developed that automatically outputs stenotic segments in visible main coronary arteries. The model consisted of multiple consecutive steps, as shown in Figure 1. The model started by leveraging a deep learning segmentation model (*nnUnet*⁽²³⁾) that outputs the segmentation masks of the three main arteries (if present)⁽¹⁰⁾. Each branch was then processed subsequently by extracting the border from the segmentation mask, which allowed the computation of the diameter profile of each branch. Finally, the diameter profile was analyzed: if the diameter was locally largely under its theoretical value, a stenosis was detected. The width, length, and severity (QCA) of the stenosis are measured, similar to the study of Liu et al.⁽²⁴⁾

Model Training and Testing

We trained several versions of the coronary artery segmentation: i) a version that was trained on data from Center 1 only (Model 1); ii) a version trained on data from Center 2 only (Model 2); iii) a version trained on data from both Center 1 and Center 2 (Model 3). To ensure a more balanced comparison, a fourth (Model 1B) model was introduced, which was trained on a subset of data from Center 1 of the same size as the training set from Center 2. The fifth model (Model 1C) was trained on a subset of data from Center 1, matched in size and number of images with stenosis to Center 2. An overview of the training and testing processes is depicted in Figure 2.

To train each of the models, we used a 5-split patient-wise strategy. The dataset was randomly divided into five equal parts: four for training and one for validation. In each training iteration a

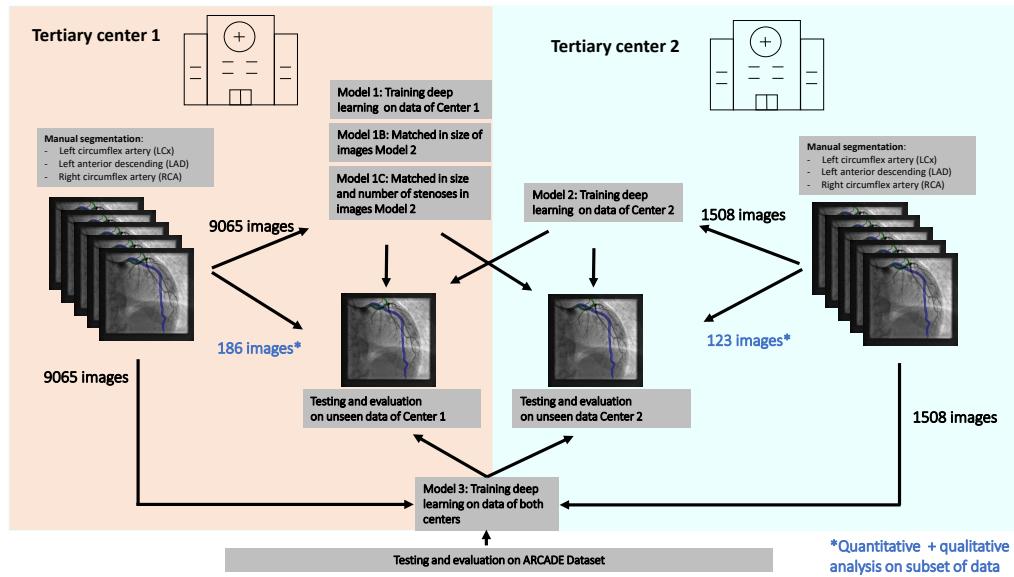


Figure 2: Acquisition of Datasets, Training and Testing. Training and testing of deep learning models for the segmentation of coronary arteries and the detection of significant stenosis on invasive coronary angiography using datasets from Center 1 (Model 1, Models 1B and 1C), Center 2 (Model 2), and combined datasets from both Center 1 and Center 2 (Model 3).

different split served as the validation set.

The code was written in Python utilizing PyTorch. The optimization ran for 150 iterations in 10 hours to 2 days (depending on the size of the training set) on a GPU Nvidia Titan RTX.

Quantitative Model Evaluation

The segmentation quality of the main arteries was assessed using the Dice coefficient, which measures the overlap between the predicted and ideal segmentations. The Dice coefficient is calculated as $2\text{TPvess} / (\text{2TPvess} + \text{FNvess} + \text{FPvess})$, where TPvess, FNvess, and FPvess represent the true positives (pixels that are present in both the predicted and gold standard), false negatives (pixels in the gold standard, not in the prediction), and false positives (pixels in the prediction, not in the gold standard), respectively. We report the Dice coefficient for the three main branches RCA, LAD, LCX separately, and a combined measurement.

A second indicator was the algorithm's ability to detect and quantify stenotic segments in the main artery branches. We designed a stenosis detection accuracy indicator, starting by selecting a single focal point at the center the narrowest segment of each stenosis identified by the expert. We then checked if this focal point fell within the predicted stenosis segmentation. In such cases, we classified the stenosis as correctly detected, incrementing TPstenosis. If the model missed a stenosis, FNstenosis incremented. The stenosis detection rate was reported and defined as the ratio of the stenosis marked by the experts that were detected by the model $\text{TPstenosis} / (\text{TPstenosis} + \text{FNstenosis})$. In addition, the false positives FPstenosis generated by the model were measured and defined as the regions predicted as stenotic but not covering any gold standard

stenosis focal point.

After training each of the models on different splits, these were applied to the test data set from both centers (comprising 49 patients from Center 1 and 32 from center 2) to evaluate the segmentation performance and the stenosis detection rate. Bootstrap samples were created with replacement, on which the models were evaluated. The segmentation performance and stenosis detection rate of these models were averaged to obtain a robust mean estimate. The 95% confidence interval (CI) for the stenosis detection rate were estimated by applying one of the five split-wise trained models to the bootstrap dataset. The 95% CI was not calculated for the combined set of all split-wise trained models to address the potential bias from shared training data.

To further evaluate the models' performance on data of an external center, the models were applied to the test dataset of the Automatic Region-based Coronary Artery Disease Diagnostics using X-ray angiography images (ARCADE) dataset⁽²⁴⁾. This publicly available dataset consists of 300 annotated ICA images, including coronary artery segmentation and stenosis. The process for applying our models to the ARCADE dataset are outlined in the Appendix.

To investigate how segmentation performance increases with the amount of data, the models were retrained on subsets of their original dataset.

6

Quantitative Inter Expert Evaluation

Along with metrics regarding the model performance, we also measured inter-expert variability. Specifically, the stenoses visible in the test dataset were reannotated by one expert from each center. As a result, each image from the test set had three sets of expert annotations: i) the original annotation from the acquisition center (comprising gold standard arteries and gold standard stenotic regions), ii) an additional annotation from Center 1 (stenoses only), and iii) an additional annotation from Center 2 (stenoses only).

Stenosis detection metrics (TPstenosis, FNstenosis and FPstenosis) were computed for the expert annotations in the same manner as for the model predictions. Each annotation served as a reference, and we counted the number of the stenoses located within the stenotic area marked by another expert, updating the measurements for TPstenosis, FNstenosis and FPstenosis. These metrics were computed to estimate the variability among expert observers by comparing annotations from experts within the same center (intra-center variability) and between centers (inter-center variability).

Qualitative Model Evaluation

Qualitative evaluation involved a visual assessment of the stenosis segmentation contours generated by Model 3. The assessment was conducted by two experienced interventional cardiologists, one from each tertiary center. The contours were assessed alongside the original ICA image on the same monitor. The quality of the segmentations was assessed by the cardiologists using a scoring system based on three statements, as shown in the Appendix and Supplemental

Table 1. These statements focused on the correctness and accuracy of the segmentation at the stenosis, graded by “yes/no” or a three-point grading scale.

Statistical Analysis

Student’s t-test for paired sample analysis was performed to investigate whether there was a significant difference in quality of the segmentations produced by the different models (Models 1, 2, 3, 1B, 1C). A χ^2 contingency test of independence was performed on the paired binary stenosis detection results. This test determines whether the differences in stenosis detection rate reported for the different models (Models 1, 2, 3, 1B, 1C) or the different experts were statistically significant. In both cases, a p-value < 0.05 was considered statistically significant.

The CLAIM (Checklist for Artificial intelligence in Medical Imaging) checklist was followed to ensure transparent reproducible reporting of the study, as shown in the Appendix.

Results

Study Population

The average age of patients from Center 1 (n=2673) was 65 ± 12 years, with 66% males, while the average age of those from Center 2 (n=456) was 56 ± 14 years, with 61% males. Cardiovascular risk factors were more often reported in patients of center 2 compared with patients of Center 1, as shown in Table 1. A total of 5572 images (61%) contained one or more significant stenosis in Center 1, and 102 images (7%) in Center 2.

Table 1: Baseline characteristics

Patient characteristics	Center 1 (n=2624)	Center 2 (n=456)	P-Value
Age (years), mean (SD)	64.6 (11.8)	56.1 (13.7)	<0.001
Male, n (%)	1719 (66.1)	307 (61.4)	0.010
Smoking, n (%)	809 (31.7)	180 (37.1)	0.022
Hypertension, n (%)	1443 (56.5)	362 (72.5)	<0.001
Diabetes mellitus, n (%)	534 (20.7)	149 (29.9)	<0.001
Dyslipidemia, n (%)	690 (27.3)	252 (50.5)	<0.001
Annotated images, n	9065	1508	
Left anterior descending, n	5618	983	
Left circumflex artery, n	5618	969	
Right coronary artery, n	2738	482	
Significant stenosis ($\geq 50\%$), n	5572	102	

Testing Data

The number of images, patients, and stenosis that were annotated in the test set is depicted in Table 2. The annotators have similarly labeled stenosis of datasets from both centers, particularly for the main arteries in quantifiable regions (Supplemental Figure 1). A total of 171 images from 49 patients were annotated in Center 1. In the first round of annotations, 136 significant stenoses (QCA ≥ 0.5) were marked by an expert from Center 1. In the second round of annotations of this

dataset, 122 significant stenosis were annotated by an expert from Center 1 and 130 significant stenoses by an expert from Center 2. A total of 109 images from 32 patients were annotated in Center 2. In the first round, 31 significant stenoses were marked. In the second round of annotations of this dataset, 52 significant stenosis were annotated by an expert from Center 1, and in 41 significant stenoses by expert from Center 2.

Table 2: Number of images and patients annotated by the two centers

	Number of patients	Number of images
Center 1: Train and validation Model 1	2624	9065
Center 1: Train and validation Model 1B	538	1504
Center 1: Train and validation Model 1C	546	1508
Center 2: Train and validation data Model 2	456	1508
Center 1: Test data	49	186
Center 2: Test data	32	123

Model 1B was trained on a subset of data from Center 1 of the same size as the training set from Center 2. Model 1C was trained on a subset of data from Center 1, matched in size and number of images with stenosis to Center 2.

Segmentation Model

Illustrative examples of the segmentation models are shown in Supplemental Figure 2. The performance of the segmentation models on test sets are outlined in Figure 3 and Supplemental Table 2. When tested on data sets from both centers, Model 3 achieved the highest median Dice coefficient on the main arteries (Dice = 0.84), closely followed by Model 1 (Dice = 0.84). Similarly, when tested on Center 1 data, Model 1 and 3 had the highest median Dice coefficient on the main arteries (Dice = 0.86). When tested on Center 2 data, Model 3 had the highest performance (Dice = 0.81) on the main arteries, followed by Model 1 (Dice = 0.81) and Model 2 (Dice = 0.77). Testing the models on data of the other center decreased their median performance compared to testing on the data from the same center they were trained on. The median Dice coefficient change is 0.05 for Model 1, 0.06 for Model 1B, 0.08 for Model 1C and 0.01 for Model 2.

When focused on the arteries separately, the segmentation models of the RCA were, on average, more accurate (Dice = 0.88-0.92) than those of the LCX artery (Dice = 0.59-0.80) and the LAD artery (Dice = 0.78-0.89, Supplemental Table 2). Similar performances were found when the segmentation models were applied to the ARCADE dataset (300 images, Supplemental Table 3), except for the LAD, which demonstrated lower performance on the ARCADE dataset (Dice = 0.76-0.84) compared to a randomly selected test set (280 images) of this study (Dice = 0.81-0.88).

A decrease in segmentation performance from Model 1 to Model 1B (matched in size of data with Model 2) and 1C (matched in size of data and stenosis with Model 2) was observed in all three testing datasets, as detailed in Figure 3. The Dice coefficients measured for Model 1 were higher than those measured for Model 2 (+0.08 for the main arteries) when applied to the all test data, as shown in Figure 3, Supplemental Table 2, and Supplemental Table 4. The difference with Model 2 is smaller when the training dataset sizes were equalized (Model 1B: +0.04, Model 1C: +0.03).

Segmentation performance improves with the amount of data, reaching a plateau at approximately 4000 training samples for the LAD and LCX, and around 2500 samples for the RCA (Supplemental Figure 3). After these points, the increase in performance diminishes.

Stenosis detection

The models were applied to the unseen test set. The detection rate (the agreement between the model and the expert) of Model 3 was 64% (95% CI 60, 68), which was comparable to the detection rate of Model 1 (66% [95% CI 62, 66], P = 0.05), as shown in Figure 4, Table 4, and Supplemental Table 5). Both models were similar to the inter-expert agreement of 65%. The inter-expert agreement (65%), inter-center agreement (64%), and intra-center agreement (65%) rates were all similar. The detection rate decreased when the model was trained on smaller datasets from one of the centers. The detection rate of Model 2 was lower (38% [95% CI 34, 42]) compared to Model 1 (P < 0.001) and Model 1B (58% [95% CI 54, 62], P < 0.001). Model 1C, which was trained with a matched size and stenosis rate to Model 2, had a comparable detection rate (Model 1C: 42% [95% CI 38, 46], P = 0.10). The relative performance of the different models varied depending on whether they were trained and tested at the same Center. The mean difference in stenosis detection rate between Model 1B (58%) and Model 2 (38%) was 20% on the complete test set, 21% (Model 1B: 56%, Model 2: 35%), on the Center 1 test set, and 13% (Model 1B: 69%, Model 2: 56%) on Center 2 test set (Table 3). The assessment of stenoses with a measured QCA ≥ 0.5 and QCA > 0.7 increased the inter-expert agreement to 68% for both thresholds. For both cut-off points the stenoses detection rates for the models increased compared to the detection rates for all stenoses assessed, as depicted in Supplemental Figure 4.

Table 3: Stenosis Detection rates for quantifiable stenoses in the main arteries with 95% confidence intervals

	Model 1	Model 1B	Model 1C	Model 2	Model 3
All test data	66% [62%, 70%]	58% [54%, 62%]	42% [38%, 46%]	38% [34%, 42%]	64% [60%, 68%]
Test data Center 1	64% [60%, 68%]	56% [51%, 61%]	39% [34%, 43%]	35% [30%, 39%]	62% [57%, 66%]
Test data Center 2	72% [64%, 80%]	69% [60%, 77%]	57% [49%, 66%]	56% [47%, 65%]	73% [56%, 80%]

Table 4: Stenosis detection rate for significant lesions in the main arteries using all test data

	RCA	LAD	LCX	All main arteries
Model 1	75% [70%, 80%]	69% [63%, 75%]	52% [49%, 55%]	66% [62%, 70%]
Model 1B	67% [60%, 74%]	60% [54%, 66%]	46% [41%, 51%]	58% [54%, 62%]
Model 1C	47% [39%, 55%]	49% [44%, 54%]	25% [16%, 34%]	42% [38%, 46%]
Model 2	47% [39%, 55%]	41% [36%, 46%]	23% [12%, 34%]	38% [34%, 42%]
Model 3	71% [67%, 75%]	66% [62%, 70%]	52% [46%, 58%]	64% [60%, 68%]

The stenosis detection rate for the models is highest for the RCA (47%-75%), followed by the LAD (41%-69%), and then the LCX (23%-52%), as shown in Table 4. The detection rate for stenoses in the LCX is low for Model 1C (25% [16%, 34%]) and Model 2 (23% [12%, 34%]). The mean detection rate for LCX stenoses is nearly twice as high with model 1B (46% [41%, 51%]).

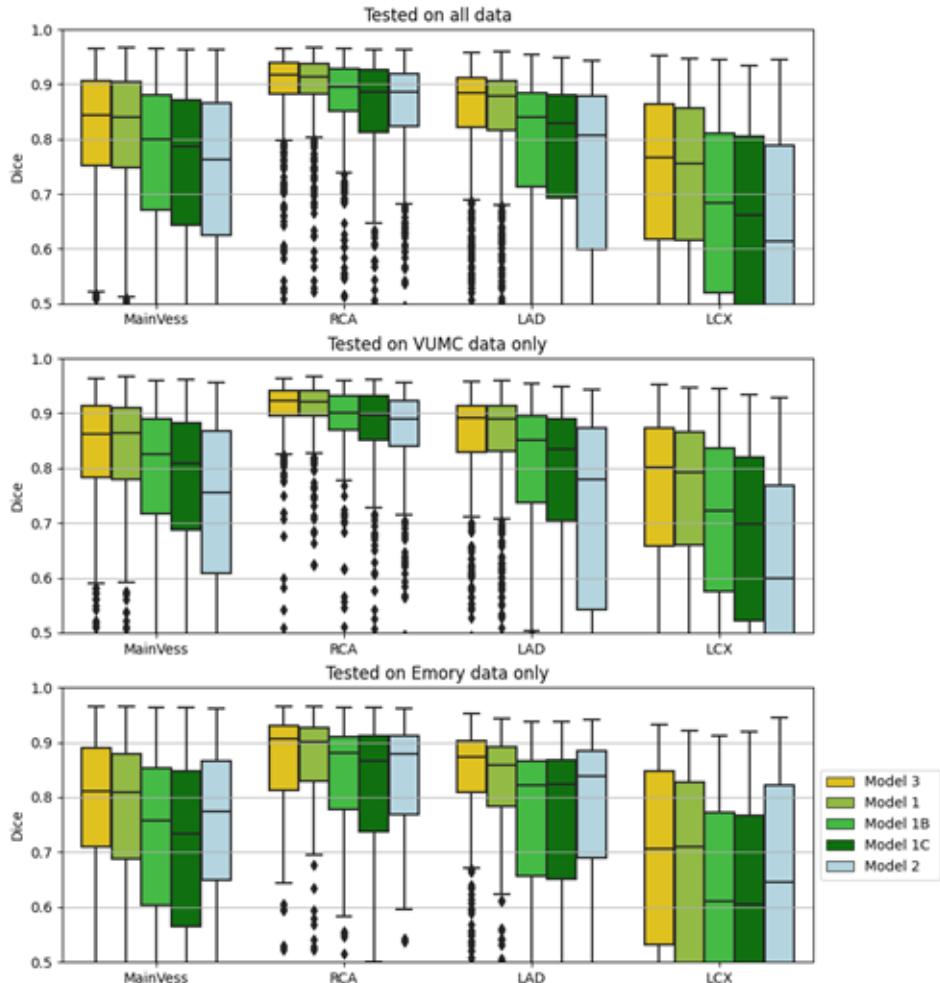


Figure 3: Evaluation of Segmentation Models on Unseen Test Data. The segmentation models were tested on all test data (top plot), on test data from Center 1 only (middle plot), and on test data from Center 2 only (bottom plot). The evaluation metric used was the Dice coefficient. The segmentation model trained on data from both centers (Model 3) achieved the highest average Dice coefficient compared to the other models for annotating the main arteries combined (MainVess), and the individual arteries (RCA, LAD, LCX).

Specifically, the detection rates on the distal segments are lower compared to mid and proximal segments, as shown in Supplemental Table 6.

The mean values for correct predictions (TPstenosis) and false predictions (FPstenosis) for stenosis detection per image are presented in Figure 5. The TPstenosis and FPstenosis were calculated per image for expert agreement (TPstenosis 0.67, FPstenosis 0.40). Model 1 (TPstenosis 0.66, FPstenosis 0.49) and Model 3 (TPstenosis 0.63, FPstenosis 0.44) showed slightly higher true positive rates compared to Model 1B (TPstenosis 0.59), Model 1C (TPstenosis 0.42), and Model 2 (TPstenosis 0.38). Model 1 and Model 3 achieved the best result when applied to the ARCADE dataset, as shown in Supplemental Table 7. The TP ratio, defined as the number of true positive stenosis divided by the total number of stenosis in the main arteries, for Model 1B (64%) was similar to that of Model 1 (64%) and Model 3 (64%), but showed 20% more false positives per image. The TP ratio of model 1C was the lowest at 45% (with 0.55 false positives per image), followed by Model 2 at 49% (with only 0.26 false positives per image).

Qualitative analysis

The segmentation contours generated by Model 3 were qualitatively evaluated using 55 angiograms of 16 patients. Experts from Center 1 and Center 2 assessed that 72% and 74% of the stenoses in the main arteries were detected by the model, respectively, as shown in Table 5 and Supplemental Table 8. Among all stenoses detected by Model 3, 69% were identified as significant stenoses by Expert 1, and 83% by Expert 2. The two experts agreed on the significance in 73% of the detected stenoses. Stenosis length measurements were classified as accurate or with minor error in 76% of cases by Expert 1 and 95% of cases by Expert 2. The minimal stenosis diameters were classified as accurate or with minor error in 82% of the cases by Expert 1 and 90% of the cases by Expert 2.

Table 5: Qualitative analysis of 55 angiograms by two interventional cardiologists

Variable	Expert 1 (VUMC)	Expert 2 (Emory)
Stenosis detection		
True positive stenosis	33	40
False positive stenosis	15	8
True negative stenosis	15	14
False negative stenosis	13	14
Stenosis detection rate / True positive rate (TP/TP+FN)	72%	74%
Positive Predictive Value (TP/TP+FP)	69%	83%
True negative rate (TN/TN+FP)	50%	64%
False positive rate (FP/FP+TN)	50%	36%
F1 score (2*TP/(2*TP+FP+FN))	70%	78%
Stenosis characteristics		
Percentage of stenosis with precise length or minimal error	76%	95%
Percentage of stenosis with precise diameter or minimal error	82%	90%

FN = False Negative, FP = False Positive, TP = True Positive, TN = True Negative.

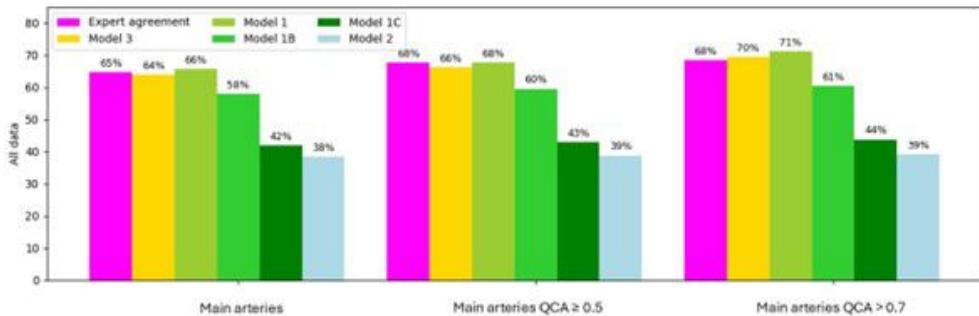


Figure 4: Evaluation of Stenosis Detection Rate on Unseen Test Data. The stenosis detection accuracy has been computed for: quantifiable stenoses located on the main arteries (left plot), quantifiable stenoses located on the main arteries with a $QCA \geq 0.5$ (middle plot), and quantifiable stenoses located on the main arteries with a $QCA > 0.7$ (right plot).

6

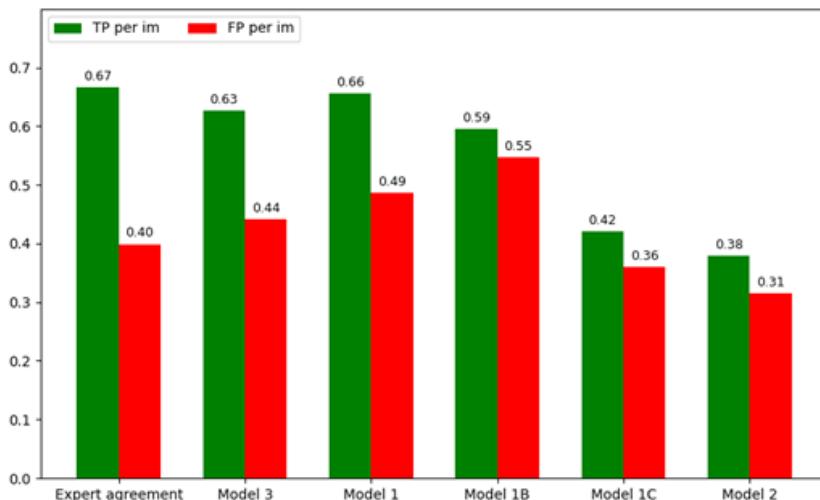


Figure 5: Evaluation of Stenosis Prediction Performance on Unseen Test Data. Mean True Positives (TP) and false positives (FP) of stenosis detection per image were calculated for the five tested models, compared with expert variability.

Discussion

This study evaluated the performance of deep learning models for the segmentation of coronary arteries and the detection of significant stenosis on ICA. We demonstrated that (i) there is a decrease in performance when models are tested on data from other centers; (ii) not only the quantity of training data drives the quality of the model, but also some characteristics of this data, such as the number of cases with significant stenosis; and (iii) when trained on an extensive and representative dataset, the models perform on par with experts for the detection of stenoses.

A reduction in performance when testing a model at another center is not uncommon, but it has not been explored as extensively as in this study by reciprocal testing across centers. Most previous studies on automated coronary angiography analysis perform validation on a subset of the data that is not used for training. This dataset is often annotated within the same center by the same annotators. Du et al.⁽²¹⁾ trained a deep-learning model, called DeepDiscern, to automatically segment the coronary arteries and detect lesions on a large data set (segmentation model: 12,323 images, lesion detection: 6,239 images). On an unseen test set, the segmentation algorithm demonstrated an accuracy of 0.98 and a sensitivity of 0.85. The average lesion detection rate was 0.90. Although DeepDiscern showed good performance, the model was not evaluated on data from another center⁽²¹⁾. If studies perform external validation, they typically validate their models on one dataset, which is often small in size compared to the dataset the model was trained on. Yang et al.⁽¹⁰⁾ trained a deep-learning model on 3302 angiographic images (annotated by two experts) and demonstrated a slight reduction in segmentation performance when validated in an external center (181 images), with a decrease in Dice from – 0.02 (from 0.92 to 0.90). In our study, the decrease in segmentation performance for Model 2 was minimal (-0.01), while it was larger for Model 1 (-0.05) and Model 1C (-0.08). The ICA images of Center 1 were annotated by ten experts, which may have led to inconsistencies in the annotations (e.g. varying delineations of arteries and distal artery segmentation) and thereby the model's performance. For example, in our results, there was considerable variability in the performance of the models trained for segmenting the LCX artery. This might be explained by the variability in how the LCX was drawn distally by the annotators. The model may not have learned a consistent representation of the distal LCX. These findings are in line with the study of Yang et al.⁽¹⁰⁾, which similarly observed a noticeable decrease in performance of the LCX segmentation models when tested in an external center. The segmentation models of the RCA were accurate (Dice = 0.88–0.92), which may be explained by the relatively simple appearance and consistency in how the RCA is segmented.

Several other studies have focused on various components of the pipeline required for the automated analysis of coronary angiography. The following components have been described: video and frame selection^(9, 25, 26), anatomy localization or segmentation^(9, 10, 21), stenosis detection^(9, 21, 27), and stenosis assessment⁽⁹⁾. In recent years, deep learning models have played an increasingly important role in automating these components of coronary angiography analysis. The deep-learning applications for segmentation used in this and previous studies offer

advantages over other segmentation methods, such as tracking⁽²⁸⁾, model⁽²⁹⁾, or filter-based models, in terms of processing steps and accuracy⁽²⁹⁾.

Recently, Avram et al.⁽⁹⁾ developed a pipeline, called CathAI, which addresses all components of automated coronary angiography interpretation using a series of algorithms. However, the segmentation algorithm showed suboptimal performance for identifying coronary artery segments and was only able to detect the coronary artery segments as bounding boxes. Furthermore, the stenosis localization was poor (14% average precision for the model trained on the left and right coronary arteries, and 26% for the model trained on the RCA). A follow-up of CathAI, DeepCoro⁽³⁰⁾, further improves the identification of coronary artery segments by refining the delineation of the artery and focusing on video-based deep learning algorithms in which sequences of frames are processed. DeepCoro was evaluated on 11 specific coronary segments of 200 images and achieved an average Dice coefficient of 0.74 for segmenting the coronary arteries, which was lower compared with the values found in our study (Dice coefficients in internal validation Model 1: 0.86; Model 2: 0.77; Model 3: 0.84). Larger training sets in our study (Center 1: 9065 images, Center 2: 1508) and a segmentation method with focus on main arteries (not on segments) may account for the difference in performance. The DeepCoro stenosis detection algorithm (trained on 1335 videos including 300 severe stenoses) demonstrated a detection rate (sensitivity) for \geq 70% stenosis of 67% (image-based model) and 73% (video-based model) when applied to 333 videos (with 71 severe stenoses). These detection rates were comparable to Model 1 (71%) and Model 3 (70%), but not to Model 1C (44%) and Model 2 (39%), which were trained on datasets with a low number of stenoses (7%). The application of the models to the ARCADE dataset showed that Model 1C and Model 2 had low TP ratios compared to the models trained on datasets with a larger number of stenoses. These findings suggest that it is important to train a model on a dataset with a substantial size and number of stenosis to achieve good performance when applied to other centers. In addition, we have observed that combining data of multiple centers (Model 3) seems to further improve performance compared to using data from one center. These factors are important to consider when training and testing these models, especially as further work progresses towards automated QCA, non-invasive fractional flow reserve estimation⁽³¹⁾, and potentially automated SYNTAX scoring^(11, 21).

Previous studies have excluded patients for a wide range of reasons, including chronic total occlusion (CTO)^(9, 32), percutaneous coronary intervention⁽⁹⁾, severe artery overlap⁽³²⁾, diffuse stenosis⁽¹⁸⁾, or patients without a stenosis^(10, 32), or have focused on training models exclusively for the right coronary artery⁽³³⁾. These factors complicate comparisons between studies and hinder the application of these models to real-world patients. Another important factor is that studies often lack validation in an external center^(21, 27, 34, 35). The performance of these models is probably lower in another center, as demonstrated in this study. If models are evaluated in another center, it is important to evaluate models on datasets that have been annotated in a similar manner. For example, out-of-plane mid and distal segments of the LAD were not all segmented in the ARCADE dataset, which required modifications to the evaluation process, as detailed in the Appendix.

If validation is performed, it is often performed against a single expert, which does not account for the inter-user variability that exists when interpreting coronary angiography images^(6, 8). A strength of this study is that we compared the stenosis detection models to expert agreement. We observed that the mean number of false positive stenosis predictions per image of these models were slightly higher (Model 1 0.49, Model 3 0.44) compared to experts (0.40), while the mean number of true positive stenosis predictions per image (Model 1 0.66, Model 3 0.63) was comparable to expert agreement (0.67).

In quantitative analysis, there was a small difference in the number of annotated images in the test set between the first and second round of experts. This may be caused by different annotator sensibilities regarding which stenoses are significant and which are not. Only preselected images were interpreted, and the runs were not provided to the experts. This approach has the advantage of presenting the experts with the precise set of information that the algorithm utilizes. However, it makes it difficult to robustly assess whether there is a significant stenosis, whether there is a single stenosis or multiple separate stenoses, and whether a stenosis is quantifiable. This might explain the relatively low agreement between observers. However, the agreements found in quantitative (65%) and qualitative analyses (73%) align with previous studies that have reported an expert agreement in the significance of stenosis in 65%⁽⁶⁾ and 81%⁽⁸⁾ of the coronary angiograms. These numbers emphasize the importance of developing models to reduce variability in clinical interpretation. The application of these models as independent observers, assisting physicians in the interpretation of ICA images, could enhance the reliability and accuracy of stenosis assessments. This improvement may lead to better-informed clinical decisions and potentially improve patient outcomes⁽³⁰⁾.

Several remarks should be made about this study. First, the models trained were only restricted to the main arteries. Smaller branches of the coronary arteries were not considered. Second, qualitative assessment provided additional information about the stenosis detection algorithm trained on data of both centers. The lengths and diameters of the stenosis were assessed as accurate or with minor errors in most cases, which suggest that the algorithm can help provide physicians with QCA information. Third, speed of the algorithms and time investment required from physicians is increasingly important when implementing novel digital solutions⁽³⁶⁾, especially when algorithms need to perform in real-time. In this study, the processing of single images was fast and typically ranged from 2 to 7 seconds. These values were in line with the Deep-Discern study, which recorded 1.3 seconds for segment recognition and 0.7 seconds for lesion detection⁽²¹⁾. However, analyzing multiple frames, as is required when processing ICA videos, will take longer. This was shown by the DeepCoro pipeline that had an average processing time of 63 seconds to generate predictions for a video⁽³⁰⁾. Fourth, training and testing was performed on data of two centers; the models were further evaluated on the ARCADE dataset⁽²⁴⁾. It is further expected that including data from more centers for training the models will continue to improve the performance. This may result in more standardized and reliable assessment of stenoses by serving as an independent observer in the ICA interpretation, and may lead to a higher degree of diagnostic accuracy. Fifth, all images were acquired using Philips equipment, which limits the

generalizability to centers that use imaging systems from other vendors. However, we evaluated the models on the ARCADE dataset, which include images acquired using Philips and Siemens imaging systems. Our analysis showed that, except for the segmentation of the LAD, there were no significant differences in performance compared with our test set. It remains challenging to determine whether the decline in LAD performance was due to the difference in ground truth annotation or an actual degradation of the model performance on data from another center. Sixth, using video-based models have the potential to improve performance over image-based models used in this study. The temporal information video-based models use is necessary to understand the dynamics of the cardiovascular system, mimicking how physicians analyze ICA videos⁽³⁰⁾.

Conclusion

Among the multiple deep learning models trained, two models demonstrated a performance on par with experts for the segmentation of coronary arteries and the detection of significant stenosis in the main arteries. The findings of this study demonstrate that dataset size, stenosis prevalence, and the training and testing center are important factors to consider when training models for automated coronary angiography analysis.

References

- [1] Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The Global Burden of Cardiovascular Diseases and Risk. *Journal of the American College of Cardiology* 2022;80:2361–2371.
- [2] Timmis A, Vardas P, Townsend N, Torbica A, Katus H, De Smedt D et al. European Society of Cardiology: cardiovascular disease statistics 2021. *European Heart Journal* 2022;43:716–799.
- [3] Lawton JS, Tamis-Holland JE, Bangalore S, Bates ER, Beckie TM, Bischoff JM, et al. 2021 ACC/AHA/SCAI Guideline for Coronary Artery Revascularization: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2022;145:e18–e114.
- [4] Kobayashi T, Hirshfeld JW. Radiation Exposure in Cardiac Catheterization. *Circulation: Cardiovascular Interventions* 2017;10:e005689.
- [5] Zhang H, Mu L, Hu S, Nallamothu BK, Lansky AJ, Xu B, et al. Comparison of Physician Visual Assessment With Quantitative Coronary Angiography in Assessment of Stenosis Severity in China. *JAMA Intern Med* 2018;178:239–247.
- [6] Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver variability in coronary angiography. *Circulation* 1976;53:627–632.
- [7] Nallamothu BK, Spertus JA, Lansky AJ, Cohen DJ, Jones PG, Kureshi F, et al. Comparison of Clinical Interpretation With Visual Assessment and Quantitative Coronary Angiography in Patients Undergoing Percutaneous Coronary Intervention in Contemporary Practice. *Circulation* 2013;127:1793–1800.
- [8] Leape LL, Park RE, Bashore TM, Harrison JK, Davidson CJ, Brook RH. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. *Am Heart J* 2000;139:106–113.
- [9] Avram R, Olglin JE, Ahmed Z, Verreault-Julien L, Wan A, Barrios J, et al. CathAI: fully automated coronary angiography interpretation and stenosis estimation. *npj Digit Med* 2023;6:1–12.
- [10] Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep* 2019;9:1–11.
- [11] Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, et al. Current State and Future Perspectives of Artificial Intelligence for Automated Coronary Angiography Imaging Analysis in Patients with Ischemic Heart Disease. *Curr Cardiol Rep* 2022;24:365–376.
- [12] Schuurings MJ, Işgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine Echocardiography and Artificial Intelligence Solutions. *Frontiers in Cardiovascular Medicine* 2021;8.
- [13] Nasr-Esfahani E, Karimi N, Jafari MH, Soroushmehr SMR, Samavi S, Nallamothu BK, et al. Segmentation of vessels in angiograms using convolutional neural networks. *Biomedical Signal Processing and Control* 2018;40:240–251.
- [14] Cervantes-Sanchez F, Cruz-Aceves I, Hernandez-Aguirre A, Hernandez-Gonzalez MA, Solorio-Meza SE. Automatic Segmentation of Coronary Arteries in X-ray Angiograms using Multiscale Analysis and Artificial Neural Networks. *Applied Sciences* 2019;9:5507.
- [15] Nobre Menezes M, Silva JL, Silva B, Rodrigues T, Guerreiro C, Guedes JP, et al. Coronary X-ray angiography segmentation using Artificial Intelligence: a multicentric validation study of a deep learning model. *Int J Cardiovasc Imaging* 2023;39:1385–1396.
- [16] Nobre Menezes M, Lourenço-Silva J, Silva B, Rodrigues O, Francisco ARG, Carrilho Ferreira P, et al. Development of deep learning segmentation models for coronary X-ray angiography: Quality assessment by a new global segmentation score and comparison with human performance. *Rev Port Cardiol* 2022;41:1011–1021.
- [17] Liang D, Qiu J, Wang L, Yin X, Xing J, Yang Z, et al. Coronary angiography video segmentation method for assisting cardiovascular disease interventional treatment. *BMC Med Imaging* 2020;20:65.
- [18] Iyer K, Najarian CP, Fattah AA, Arthurs CJ, Soroushmehr SMR, Subban V, et al. AngioNet: A Convolutional Neural Network for Vessel Segmentation in X-ray Angiography
- [19] Gao Z, Wang L, Soroushmehr R, Wood A, Gryak J, Nallamothu B, et al. Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features. *BMC Medical Imaging* 2022;22:10.
- [20] Wang L, Liang D, Yin X, Qiu J, Yang Z, Xing J, et al. Coronary artery segmentation in angiographic videos utilizing spatial-temporal information. *BMC Medical Imaging* 2020;20:110.
- [21] Du T, Xie L, Zhang H, Liu X, Wang X, Chen D, et al. Training and validation of a deep learning architecture for the automatic analysis of coronary angiography. *EuroIntervention* 2021;17:32–40.
- [22] Serruys PW, Morice M-C, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *N Engl J Med* 2009;360:961–972.
- [23] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–211.
- [24] Popov M, Amanturdieva A, Zhaksylyk N, Alkanov A, Saniyazbekov A, Aimyshev T, et al. Dataset for Automatic Region-based Coronary Artery Disease Diagnostics Using X-Ray Angiography Images. *Sci Data* 2024;11:20.
- [25] Ciudel C, Turcea A, Puia A, Itu L, Calmac L, Weiss E, et al. Deep neural networks for ECG-free cardiac phase and end-diastolic frame detection on coronary angiographies. *Computerized Medical Imaging and Graphics* 2020;84:101749.
- [26] Wu W, Zhang J, Xie H, Zhao Y, Zhang S, Gu L. Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. *Computers in Biology and Medicine* 2020;118:103657.
- [27] Danilov VV, Klyshnikov KY, Gerget OM, Kutikhin AG, Ganyukov VI, Frangi AF, et al. Real-time coronary artery stenosis detection based on modern neural networks. *Sci Rep* 2021;11:7582.
- [28] Shoujun Z, Jian Y, Yongtian W, Wufan C. Automatic segmentation of coronary angiograms based on fuzzy inferring and probabilistic

- tracking. *BioMedical Engineering OnLine* 2010;9:40.
- [29] Kerkeni A, Benabdallah A, Manzanares A, Bedoui MH. A coronary artery segmentation method based on multiscale analysis and region growing. *Computerized Medical Imaging and Graphics* 2016;48:49–61.
- [30] Labrecque Langlais É, Corbin D, Tastet O, Hayek A, Doolub G, Mrad S, Tardif J-C, et al. Evaluation of stenoses using AI video models applied to coronary angiography. *npj Digit Med* 2024;7:1–13.
- [31] Arefinia F, Aria M, Rabiei R, Hosseini A, Ghaemian A, Roshampoor A. Non-invasive fractional flow reserve estimation using deep learning on intermediate left anterior descending coronary artery lesion angiography images. *Sci Rep* 2024;14:1818.
- [32] In Kim Y, Roh J-H, Kweon J, Kwon H, Chae J, Park K, et al. Artificial intelligence-based quantitative coronary angiography of major vessels using deep-learning. *Int J Cardiol* 2024;405:131945.
- [33] Moon JH, Lee DY, Cha WC, Chung MJ, Lee K-S, Cho BH, et al. Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Comput Methods Programs Biomed* 2021;198:105819.
- [34] Molenaar MA, Selder JL, Schmidt AF, Asselbergs FW, Nieuwendijk JD, Dalzen B van, et al. Validation of machine learning-based risk stratification scores for patients with acute coronary syndrome treated with percutaneous coronary intervention. *European Heart Journal - Digital Health* 2024;5:702–711.
- [35] Molenaar MA, Bouma BJ, Asselbergs FW, Verouden NJ, Selder JL, Chamuleau SAJ, et al. Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome. *European Heart Journal - Digital Health* 2024;5:170–182.
- [36] Man JP, Koole MAC, Meregalli PG, Handoko ML, Stienen S, Lange FJ de, et al. Digital consults in heart failure care: a randomized controlled trial. *Nat Med* 2024;30:2907–2913.

Supplementary material

Qualitative assessment

The following statements were assessed by the two interventional cardiologists for each stenosis in the image:

1. The significant stenosis was detected by the stenosis detection model (Model 3).
2. The stenosis in the imaging plane is accurately segmented along the length of stenosis (from proximal to distal).
3. The stenosis in the imaging plane is accurately segmented at the minimum lumen diameter.

Statement 1 was responded to with “yes” or “no”. Statements 2 and 3 were graded on a three-point grading scale, as shown in Supplemental Table 1.

CLAIM checklist

The completed CLAIM checklist is available in the online version of the published article: Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020; 2(2):e200029.

ARCADE dataset

Segmentation

To further evaluate the models’ performance on data of an external center, the models were applied to the test dataset of the Automatic Region-based Coronary Artery Disease Diagnostics using X-ray angiography images (ARCADE) dataset²⁴. This publicly available dataset consists of 300 annotated ICA images, including coronary artery segmentation and stenosis. To apply the models we had to overcome differences in the segmentation process between this study and the ARCADE dataset:

1. The artery segmentation challenge aims to generate the complete SYNTAX segmentation of the coronary arteries into 15 segments, whereas we focus only on segmenting the three main arteries: RCA, LAD, and LCX. Therefore, we have adapted the ARCADE dataset to fit our simpler class system.
2. In the ARCADE SYNTAX challenge, experts annotated only the arteries visible on the angiogram that are clinically relevant. In contrast, we have trained our models to always segment the main arteries when they are visible in the images.

Supplemental Figure 5 presents an example of the ARCADE ground truth (adapted to our three classes), compared to our model’s segmentations. The ground truth only includes annotations for the proximal LAD, while our models also covers the mid and distal segments. As a result, we believe the ARCADE syntax dataset can be leveraged for evaluation with minor adaptations: Dice coefficients are computed only for arteries present in the ground

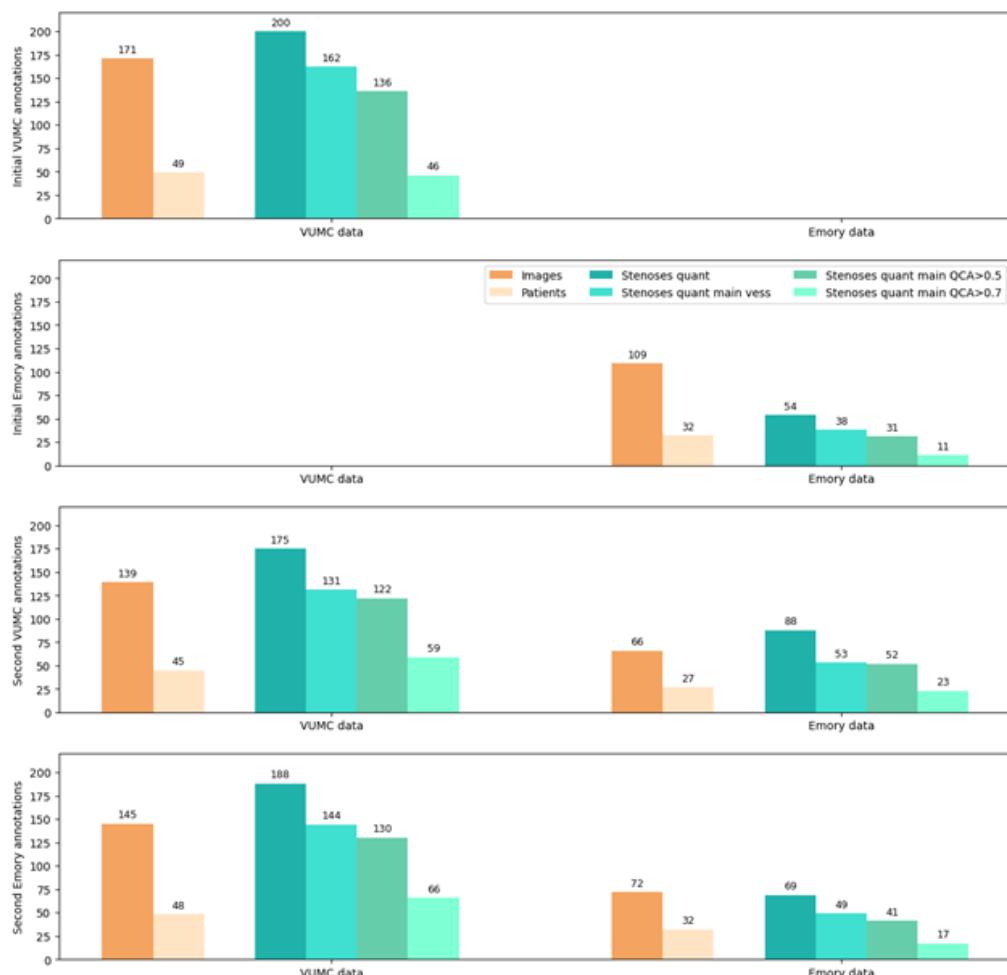
truth, and the LAD Dice coefficient is calculated only when LAD mid or distal segments are included. The cases in which only the proximal segment of the LAD was segmented are excluded.

Stenosis detection

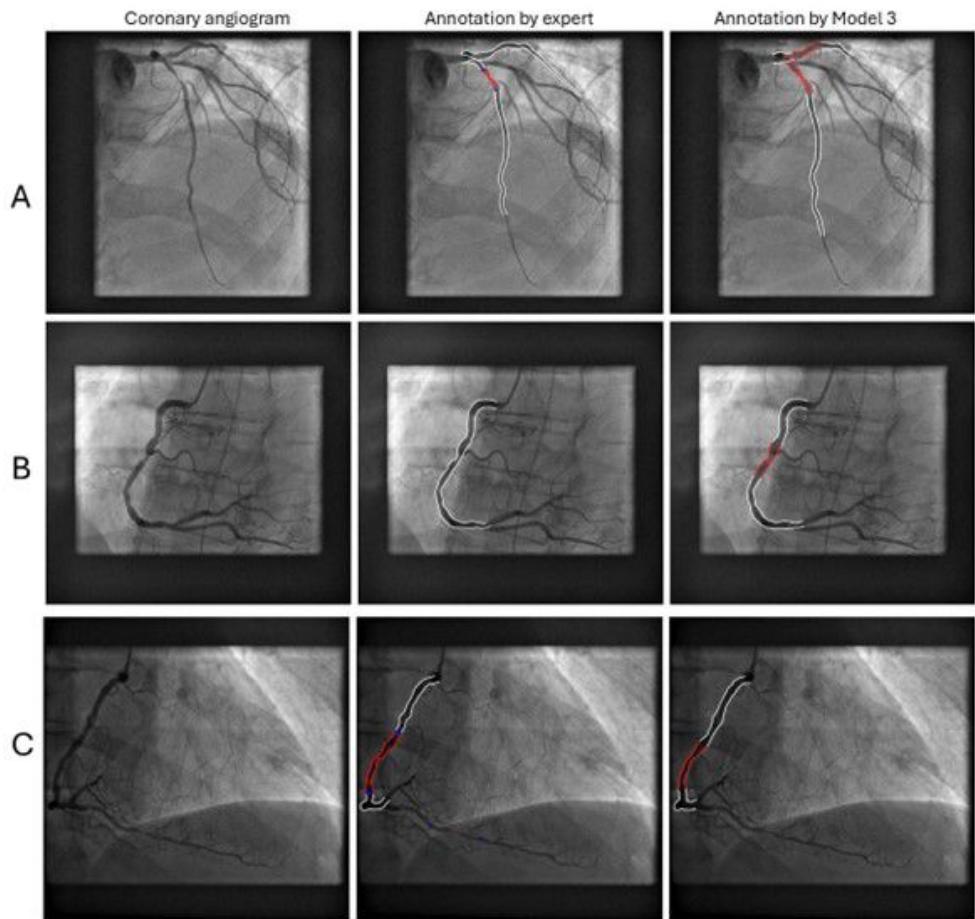
The details of the ARCADE stenosis problem statement also differ from this study. We focus on a detection problem, measuring True Positives (TP) as ground truth focal points that fall within our predicted regions. However, the ARCADE challenge emphasizes stenosis segmentation, where the ground truth is provided as segmented regions with less focus on precisely aligning with artery borders. We evaluate our models on this dataset, applying the following adaptations:

1. In this study, we focus only on stenoses located in the main arteries, defining them as regions with at least 25% overlap with the model's main artery segmentations. After this step, 17% of the stenoses marked in the ARCADE database are excluded.
2. A stenosis is considered correctly detected if it intersects a predicted region (TP). A False Positive (FP) is defined as a predicted stenotic region that does not intersect any ground truth stenotic regions.

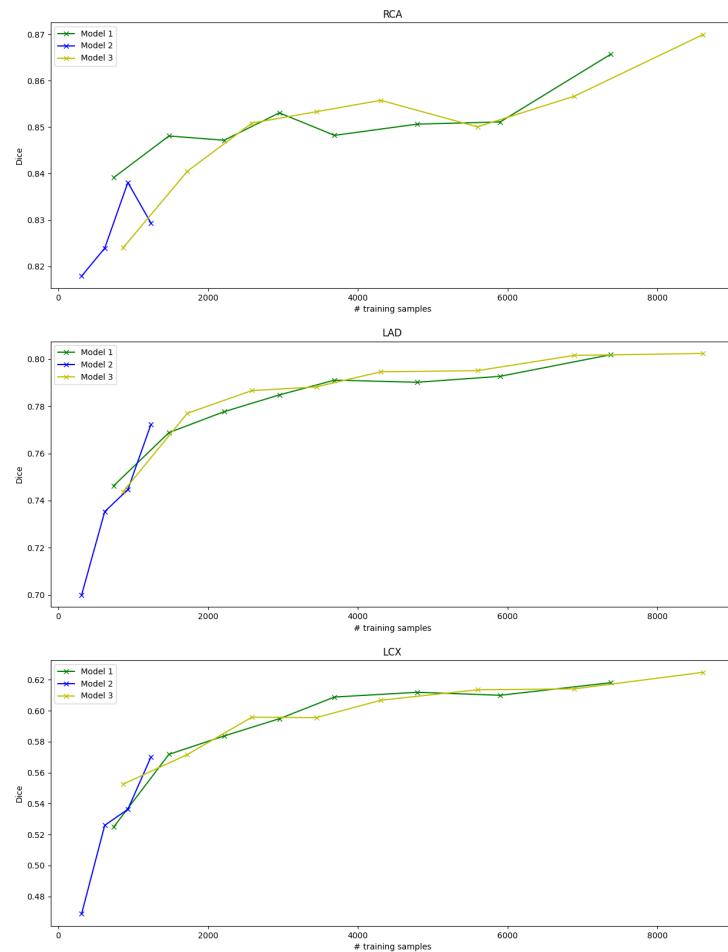
Supplemental Figure 6 presents an example of the ARCADE ground truth and the prediction by Model 3. For comparison, we applied the models to a randomly selected set of images (from the test set) similar in size to the ARCADE dataset (280 images).



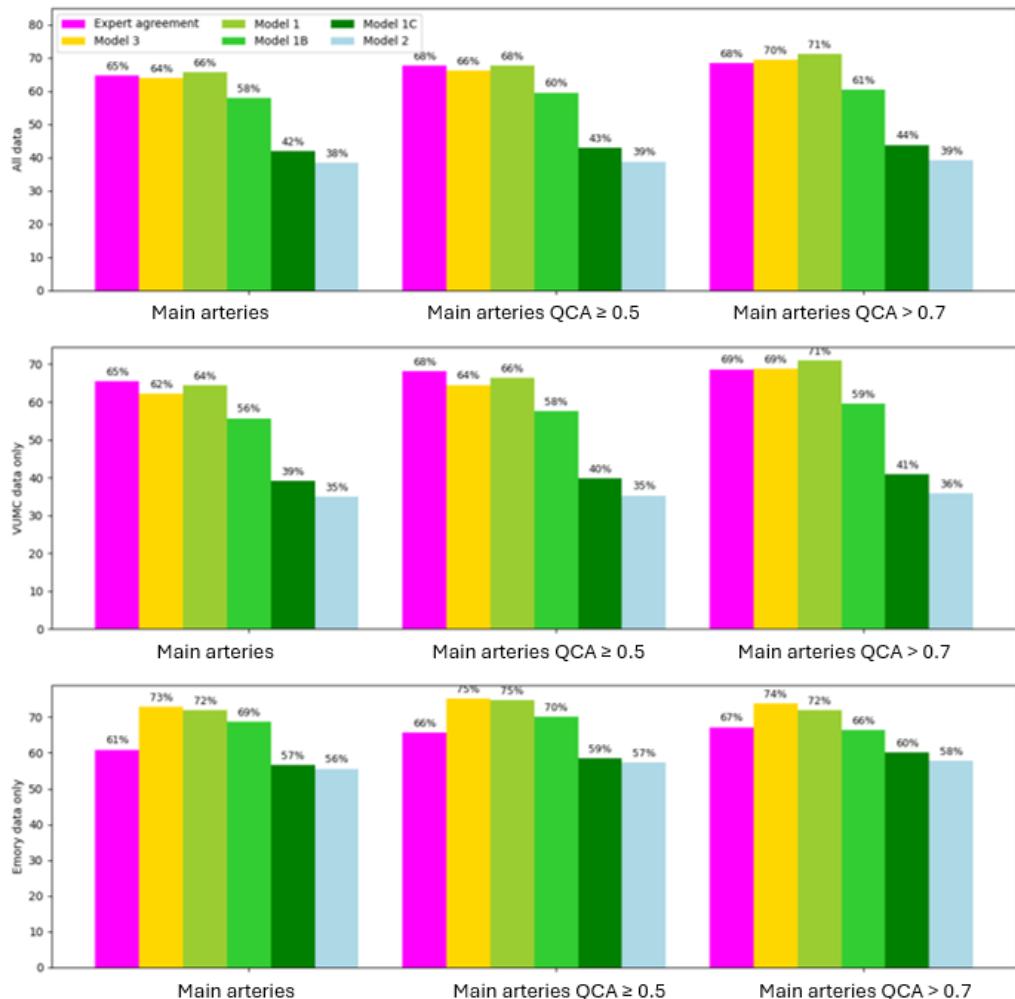
Supplementary Figure 1: Evaluation of Homogeneity in Stenosis Detection of the Experts from the Two Centers. The test sets from Center 1 (VUMC) and Center 2 (Emory) were annotated by experts from their centers, as depicted in plots A and B. The test sets from Center 1 and Center 2 were re-annotated by experts from both centers, as shown in plots C and D. The number of images and patients considered in the second rounds of annotations (orange bars) is lower in plots C and D compared to plots A and B. This reduction is due to the fact that only images and patients with stenoses were counted during the second rounds, while those with non-stenosed arteries were ignored.



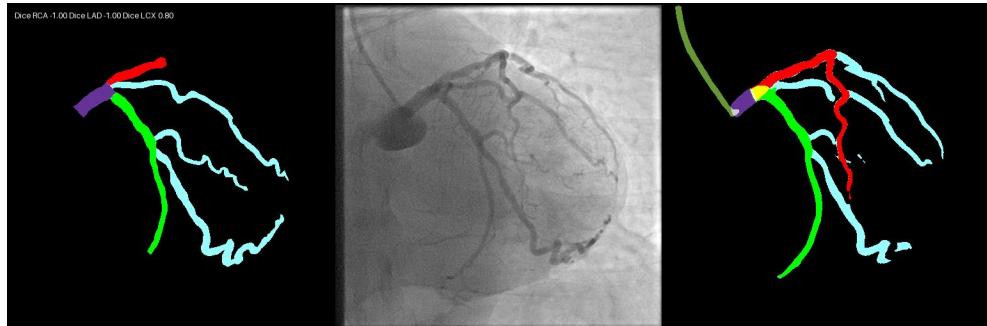
Supplementary Figure 2: Illustrative Examples of Automated Segmentation of the Coronary Arteries and Lesion Detection on Coronary Angiography Images. In Examples A and B, model 3 misclassified coronary segments as stenosed ($\geq 50\%$). In Example C, the segmentation and lesion detection align with expert annotation.



Supplementary Figure 3: Effect of Training Sample Size on Segmentation Performance of the RCA, LAD and LCX.



Supplementary Figure 4: Detection Accuracy of the Different Models. The stenosis detection accuracy has been computed for: quantifiable stenoses located on the main arteries (left plot), quantifiable stenoses located on the main arteries with a $QCA \geq 0.5$ (middle plot), and quantifiable stenoses located on the main arteries with a $QCA > 0.7$ (right plot), on all data, Center 1 data (VUMC data), and Center 2 data (Emory data).



Supplementary Figure 5: Example of artery segmentation ground truth (left) from the ARCADE dataset and the corresponding prediction by Model 3 (right). The LAD is colored red, the LCX green, and the rest of the labeled vasculature cyan.



Supplementary Figure 6: Example of stenosis detection ground truth (left) from the ARCADE dataset and the corresponding prediction by Model 3 (right). Left: Stenoses outside the main arteries are colored in gray, and those in the main arteries are colored in white. Center: Coronary angiography image. Right: Predicted main artery segmentations and detected stenoses. In this case, we would report 1 True Positive and 1 False Negative.

Supplementary Table 1: Grading matrix for qualitative evaluation

Grade	Description	Correction needed
1	Substantial errors, clinically relevant errors	Segmentation necessitates substantial correction
2	Minor error, but no clinically relevant segmentation errors	Segmentation necessitates minor correction
3	Accurate segmentation, no clinically relevant segmentation errors	No correction is needed

Supplementary Table 2: Dice coefficient for segmentation models

Dataset / Artery	Model 1	Model 1B	Model 1C	Model 2	Model 3
All test data					
Main arteries	0.84	0.80	0.79	0.76	0.84
RCA	0.91	0.90	0.89	0.89	0.92
LAD	0.88	0.84	0.83	0.81	0.88
LCX	0.75	0.68	0.66	0.61	0.77
Test set of Center 1					
Main arteries	0.86	0.82	0.81	0.76	0.86
RCA	0.92	0.90	0.90	0.88	0.92
LAD	0.89	0.85	0.83	0.78	0.89
LCX	0.79	0.72	0.70	0.59	0.80
Test set of Center 2					
Main arteries	0.81	0.76	0.73	0.77	0.81
RCA	0.90	0.88	0.87	0.88	0.91
LAD	0.86	0.82	0.82	0.84	0.87
LCX	0.71	0.61	0.61	0.64	0.71

Median values are shown. LAD= Left Anterior Descending, LCX= Left Circumflex, RCA = Right Coronary Artery.

6

Supplementary Table 3: Dice Coefficients for Segmentation Models on ARCADE Test Set and Study Test Set

Model	ARCADE Syntax (300 images)			Study Test Set (280 images)		
	RCA	LAD	LCX	RCA	LAD	LCX
Model 1	0.89	0.83	0.75	0.91	0.88	0.75
Model 1B	0.87	0.79	0.71	0.89	0.83	0.66
Model 1C	0.85	0.77	0.62	0.89	0.83	0.66
Model 2	0.87	0.76	0.74	0.89	0.81	0.61
Model 3	0.89	0.84	0.80	0.92	0.88	0.77

LAD= Left anterior descending, LCX= Left circumflex, RCA = Right Coronary Artery.

Supplementary Table 4: P-values for the difference in Dice coefficient between the different trained models

Comparison	Main artery	RCA	LAD	LCX
Model 3 vs. Model 1	0.92	0.97	0.65	0.54
Model 3 vs. Model 1B	<0.001	0.02	<0.001	<0.001
Model 3 vs. Model 1C	<0.001	<0.001	<0.001	<0.001
Model 3 vs. Model 2	<0.001	<0.001	<0.001	<0.001
Model 1 vs. Model 1B	<0.001	0.32	<0.001	<0.001
Model 1 vs. Model 1C	<0.001	<0.001	<0.001	<0.001
Model 1 vs. Model 2	<0.001	<0.001	<0.001	<0.001
Model 1B vs. Model 1C	<0.001	0.002	0.003	0.001
Model 1B vs. Model 2	<0.001	<0.001	<0.001	<0.001
Model 1C vs. Model 2	0.16	0.10	0.05	0.007

LAD= Left anterior descending, LCX= Left circumflex, RCA = Right Coronary Artery.

Supplementary Table 5: P-values for the difference in stenosis detection rates for quantifiable stenoses in the main arteries

Comparison	P-value
Model 3 vs. Model 1	0.05
Model 3 vs. Model 1B	<0.001
Model 3 vs. Model 1C	<0.001
Model 3 vs. Model 2	<0.001
Model 1 vs. Model 1B	<0.001
Model 1 vs. Model 1C	<0.001
Model 1 vs. Model 2	<0.001
Model 1B vs. Model 2	<0.001
Model 1B vs. Model 1C	<0.001
Model 1C vs. Model 2	0.10
Experts vs. Model 3	0.38
Experts vs. Model 1	0.07
Experts vs. Model 1B	<0.001
Experts vs. Model 1C	<0.01
Experts vs. Model 2	<0.001

Supplementary Table 6: Stenosis Detection Rates for Quantifiable Stenoses in the Subsegments of the Main Arteries

Artery Segment	# of Stenoses	Model 1	Model 1B	Model 1C	Model 2	Model 3
RCA prox	49	93.5%	88.2%	75.1%	78.6%	90.6%
RCA mid	47	94.0%	74.0%	50.6%	48.4%	88.5%
RCA dist	46	63.5%	58.3%	30.9%	31.5%	60.9%
LAD prox	66	80.9%	72.1%	57.3%	45.8%	75.2%
LAD mid	107	81.9%	72.3%	62.6%	57.5%	78.1%
LAD dist	44	58.6%	45.9%	34.5%	31.8%	61.4%
LCX prox	67	71.0%	60.0%	29.0%	33.6%	70.4%
LCX dist	73	47.9%	43.8%	21.4%	17.5%	51.0%

The test set was annotated through three rounds: (i) the original annotation from the acquisition center (gold standard), (ii) an additional annotation from Center 1, and (iii) another from Center 2. Each image appears three times, annotated independently. Model predictions were compared against all annotations. Dist = distal, Prox = proximal, LAD = Left anterior descending, LCX = Left circumflex, RCA = Right coronary artery.

Supplementary Table 7: Stenosis Detection Performance on ARCADE Stenosis Test Set and Study Test Set

Model	ARCADE Stenosis Dataset (300 images)		Study Test Set (280 images)	
	TP ratio	FP per image	TP ratio	FP per image
Model 1	63.6%	0.31	66%	0.49
Model 1B	64.2%	0.36	58%	0.55
Model 1C	44.8%	0.55	42%	0.36
Model 2	49.2%	0.26	38%	0.31
Model 3	64.2%	0.31	64%	0.44

TP ratio = Number of true positive stenoses divided by the total number of stenoses in the main arteries. FP = False Positive.

Supplementary Table 8: Qualitative analysis of 55 angiograms by two interventional cardiologists

Variable	Expert 1 (VUMC)	Expert 2 (Emory)
Stenosis length		
Number of stenoses with accurate length	21	34
Number of stenoses with minor length error	4	4
Number of stenoses with clinically relevant length error	8	2
Percentage of stenosis with precise length	64%	85%
Percentage of stenosis with precise length or minimal error	76%	95%
Percentage of stenosis with clinically relevant length error	24%	5%
Stenosis diameter		
Number of stenoses with accurate diameter	21	34
Number of stenoses with minor diameter error	6	2
Number of stenoses with clinically relevant diameter error	6	4
Percentage of stenosis with precise diameter	64%	85%
Percentage of stenosis with precise diameter or minimal error	82%	90%
Percentage of stenosis with clinically relevant diameter error	18%	10%

7

General discussion

Risk stratification is the cornerstone for treatment decisions in patients with CAD. It assists clinicians in identifying high-risk or rising-risk patients and facilitates resource allocation by selecting those who may benefit most from more intensive care. Current risk stratification tools are available for both manifestations of CAD, which include CCS and ACS^(1, 2). However, currently available risk stratification tools of patients with CAD are constrained by a limited number of included personalized features, have limitations in performance, or are subject to inter-observer variability. Machine learning models, models that learn from data to perform certain predictions or decisions, may improve the risk stratification of patients with CAD by providing more accurate risk prediction. Machine learning is a subfield of AI, which refers to the broad concept of computer systems performing tasks that previously required human intelligence. In this thesis, the value of machine learning for risk stratification of patients with CAD was investigated, with specific focus on TTE and ICA data.

Can machine learning-based models improve the risk stratification of CCS patients utilizing both TTE and clinical data?

Current guidelines on CCS recommend a TTE in patients with CCS for the measurement of the LV function, which is considered to be one of the strongest determinants of mortality^(1, 3). However, there is more value in TTE than is currently used for risk stratification in patients with CCS, as demonstrated in **chapter 2**. In this study, valvular heart disease emerged as an important predictor of mortality in patients with CCS. Valvular heart disease was associated with an increased risk of mortality independent of LV function and other risk factors. A machine learning model trained on these echocardiography data and clinical data demonstrated an improved performance for predicting 5-year mortality compared to LV function and other traditional risk scores (**chapter 3**). These findings suggest that machine learning may enhance long-term outcome prediction and support more accurate risk stratification in patients with CCS.

From feature selection to raw TTE data

In **chapter 3**, features were selected manually. Other studies have used machine learning models that automatically learn features from raw echocardiography images during training⁽⁴⁾. These machine learning models, known as deep learning models, have a structure in the form of a neural network, which can have multiple layers. Studies have shown promising results in the automatic extraction of laboratory values⁽⁵⁾, cardiovascular risk factors⁽⁶⁾, and mortality prediction⁽⁷⁾ directly from echocardiography images. These studies emphasize the wealth of information present in the echocardiography images for risk stratification, which can be challenging to extract visually.

Explainable machine learning

The downside of the in complexity-increasing models is that they become less explainable. To provide more insight in the importance of the features for the predictions for individual patients, explainability techniques are increasingly used. Explainability is crucial to ensure that the output of

the model is understandable, trustworthy, and transparent for both clinicians and patients^(8–10). In chapter 3, we applied the SHAP (Shapley Additive exPlanations) to provide insight into the impact of the features in the machine learning model on the prediction for the individual patient. Deep learning models are generally seen as black boxes due to their complexity. Studies on automated interpretation of echocardiography using deep learning models have employed several techniques to identify areas where the model relies on for its decision. For example, an often-used model interpretation technique for images is the Gradient-weighted Class Activation Mapping (Grad-CAM)⁽⁷⁾, which provides a heatmap for a given class prediction. Although this technique has shown that it can localize anatomical structures in echocardiography images important for mortality, it does not necessarily provide new clinically useful insights⁽⁷⁾. Another limitation of these explainability techniques is that the features impact a certain prediction but are not necessarily the cause of the prediction⁽¹¹⁾. Despite these limitations, explainability techniques remain the common standard currently available to provide insight into how the model produces its output⁽¹²⁾, and they should be considered in future studies.

Machine learning for diagnostic and prognostic imaging

TTE has unique properties including non-invasiveness, low costs, absence of radiation, and portability⁽¹³⁾. It is recommended by the ESC guidelines for assessment of cardiac structures and function in patients with CCS^(1, 3). In addition, TTE contains important information for risk prediction (chapter 2). These details make the application of machine learning on TTE data relevant to a large number of CCS patients. Machine learning has been applied for risk prediction on other imaging modalities. Recently, machine learning has shown superior results to predict mortality in patients with CCS using coronary computed tomographic angiography or stress cardiac magnetic resonance compared to traditional methods^(14, 15). However, a direct comparison of machine learning scores based on TTE data to scores derived from other imaging modalities has not yet been conducted. It is therefore unknown which imaging modality most favors the use of machine learning for risk stratification. In addition, the added value of combining data from multiple imaging modalities for risk stratification in patients with CCS is unexplored. To ensure the generalizability of a machine learning model, large datasets from multiple tertiary and non-tertiary centers are required for model training to make them robust for diverse populations and hospitals.

7



Machine learning models trained on TTE and clinical data have the potential to improve risk stratification in patients with CCS by enabling more accurate risk prediction. Studies on large datasets, including raw echocardiographic images and other diagnostic modalities, are needed to further explore the potential of these models in patients with CCS.

Can machine learning-based models improve the risk stratification of patients with ACS treated with PCI?

The number of machine learning-based risk prediction scores in literature for patients with ACS is increasing⁽¹⁶⁾. These prediction scores vary in the included features, number of features, and used models. Only a minority of these scores is validated on data from an external center. In chapter 4, two machine learning-based risk scores for predicting mortality in ACS patients treated with PCI were validated on retrospective data from a tertiary center. The two models, tested on discriminative performance (ability to discriminate between patients who are likely to die and those who are not), calibration (agreement between predicted and actual observed risk), and clinical utility by discriminative curve analysis, showed variable results. The GRACE 3.0 risk score was clinically more useful for predicting one-year mortality compared to the GRACE 2.0 risk score, which is recommended by the guidelines⁽²⁾. The PRAISE risk score, calculated with 25 variables, showed limited potential for predicting one-year mortality and did not enhance risk prediction compared to the GRACE 2.0 risk score. These findings emphasize the importance of validating models on data from external centers to test their robustness, which contributes to their reliability.

Fair, useful and reliable models

In addition to improving the reliability of the models, researchers are working on methods to ensure fairness in their outcomes⁽¹⁷⁾. Fair machine learning models are models that provide accurate predictions (and treatment recommendations) for all CAD patients without discrimination based on age, race, ethnicity, or other variables⁽¹⁸⁾. Typically, models perform worse on patient populations that are underrepresented in the data on which a model is trained⁽¹⁹⁾. For example, the GRACE 2.0 risk score was only trained on 33% female non-ST-elevation ACS patients⁽²⁰⁾ and showed a lower discriminative performance in these patients, leading to an underestimation of the risk of mortality⁽²¹⁾. Although this could lead to undertreatment, it is important to realize that a biased risk prediction model outputs do not necessarily lead to unfair decisions by the treating physician⁽²²⁾. Once a model is implemented in clinical practice, studies should investigate its clinical value by paying attention to various levels of efficacy, including prognostic performance, impact on thinking and decision-making, patient outcomes, and (societal) costs and benefits^(23, 24).



Risk stratification of patients with ACS treated with PCI utilizing machine learning models shows variable performance when externally validated in other centers. Validation of newly developed machine learning models in other centers is essential for testing their generalizability. The clinical efficacy of these models should be assessed beyond their prognostic performance alone.

Can machine learning models utilizing ICA imaging improve risk stratification?

The interpretation of ICA is prone to subjectivity, especially regarding the severity of stenosis. In chapter 6, models were trained to segment coronary arteries and detect lesions automatically. These models are a preliminary step towards objective, reproducible, and automated assessment of the burden and complexity of CAD. In addition to lesion location (chapter 6), other lesion characteristics such as calcifications, vessel tortuosity, bifurcation, or trifurcation may also be automatically extracted from ICA images⁽²⁵⁾. This information can automate the calculation of a SYNTAX-like score, which is typically a time-consuming task. Such a risk score, which both combines ICA data, clinical data, and potentially other multimodal data, may inform clinicians about the risk of adverse outcomes and aid them in their decision-making (PCI versus CABG) (chapter 5). Such an automated score is especially valuable for patients with multivessel disease, for whom the calculation of the SYNTAX score is recommended by current guidelines⁽²⁶⁾.

Data collection

The collection of annotated ICA data is a tedious process, as experienced in chapter 6. Multicenter collaborations are needed to create datasets generated by a wide variety of operators, x-ray systems, data storage methods, and annotators. To facilitate the development of models and create opportunities for external validation, a research group made their annotated ICA dataset publicly available⁽²⁷⁾. To further accelerate research into machine learning-based risk stratification, additional efforts are needed to standardize the collection, storage, and sharing of data between hospitals⁽²⁸⁾. Federated learning is a technique for hospitals to collaboratively train machine learning models, without sharing (privacy-sensitive) data⁽²⁹⁾. Federated learning is in its infancy and requires the adoption of standardized data structures⁽³⁰⁾. A new common data infrastructure in combination with the needed computational resources for model training poses challenges in the adoption of federated learning, especially in hospitals with limited information technology (IT) infrastructure⁽³⁰⁾.

7

From feature selection to raw ICA data

Limited studies have published on the prediction of mortality or other endpoints directly from ICA images without extracting predefined features. Recent evidence suggests that machine-learning can help predict a future myocardial infarction directly from raw ICA images, which provide interesting opportunities for further research⁽³¹⁾.



Machine learning-based analysis of ICA offers opportunities for risk stratification by automatically extracting predictive features, enabling reproducible and automated risk scoring (SYNTAX), and integrating clinical data to enhance the prediction of adverse outcomes.

Status of machine learning-based risk stratification for patients with CAD

The actual application of machine learning for risk stratification in patients with CAD within the clinical workflow is limited (**chapter 5**). Most machine learning products that are on the market for patients with CAD are focused on the detection of cardiovascular diseases and measurements of cardiac parameters⁽³²⁾. The number of companies that offer outcome prediction models is limited. This might be explained by the need to collect large, high-quality datasets, the uncertainty about whether it is feasible to predict outcomes, and the uncertain clinical and commercial value. In addition to imaging or signal data, risk stratification also relies on clinical and demographic data, which may be documented in unstructured reports and lack standardization between hospitals. The collection of structured data from unstructured reports may be further facilitated in the near future by the ongoing development of large language models, which show promising results for extracting information from clinical reports⁽³³⁾.

Regulatory compliance and implementation

Complying with regulations for the creation of safe and ethical AI in healthcare is also a major challenge. In 2024, the AI Act⁽³⁴⁾ entered into force, which regulates the development, implementation, and use of AI in the European Union. If AI is classified as high-risk, which includes clinical decision support systems, the AI must comply with strict assessments, including meeting certain data quality, transparency, and human oversight requirements. In addition, for high-risk applications, machine learning models need to obtain a Conformité Européenne (CE) mark from an accredited company⁽³⁵⁾. The process of CE marking is time-consuming and entails varying costs.

Barriers of clinical adoption of machine learning-based risk stratification

The machine learning-based risk scores that are made available online as risk calculators, such as the GRACE 3.0 score⁽³⁶⁾ and the PRAISE score⁽³⁷⁾ (**chapter 4**), often do not grant access to the actual machine learning model behind them. This prevents their integration into the local electronic health record system and limits their use in real-time clinical setting⁽³⁸⁾.

Another important note is the trust clinicians have in these algorithms. If it is not clear to the clinician how the model reaches a certain decision, the clinician might be averse to using the model. This so called algorithmic aversion can also occur when it is unclear who is responsible (e.g. hospital, developing company, researcher, clinician) for the output of the model⁽³⁹⁾. To reduce the risk of algorithmic aversion, transparency about the original dataset on which the model was trained is needed, along with clear explanations of the features that drive a certain output.

A prediction model is rarely actionable⁽⁴⁰⁾. Although it can estimate the risk of a clinical outcome for an individual patient, it does not necessarily guide the clinician about the appropriate clinical action to take to prevent that outcome^(40, 41). These prediction models are most effective when the resulting risk stratification aid clinical decision-making. By providing insight into the

expected outcomes of two or more treatment strategies for an individual patient, the model can support clinicians and patients in selecting the most appropriate treatment. For example, by estimating the difference in mortality risk between PCI and CABG for a specific patient, the model can guide treatment decisions.

Training a model on observational data, such as the GRACE 3.0 risk score⁽³⁶⁾, PRAISE risk score⁽³⁷⁾, and the model developed in chapter 3, may result in models that are subject to confounding bias and selection bias⁽⁴¹⁾. These models learn associations between the patient characteristics and outcomes, but not necessarily causal relationships. For example, these models might infer from the data that high-risk patients should not receive invasive treatment, even though these patients might benefit the most⁽⁴⁵⁾. This bias arises from a phenomenon known as the risk-treatment paradox^(42–44), in which clinicians are hesitant to perform invasive interventions in high-risk patients due to the perceived risk of complications and poor outcomes. As a result, the models may perpetuate this bias, potentially leading to undertreatment of high-risk patients.

Risk models can be trained on observational data but require further evaluation when estimating the effects of an intervention on outcomes. Causal inference is needed to estimate patient outcomes under different treatments by estimating causal effects, which help determine the most beneficial treatment for each risk group^(40, 46). Randomized controlled trials (RCT) are typically used for this purpose^(41, 45). For example, by randomly assigning clinicians to have access to the risk model, the average effect of using the model on patient outcomes can be estimated⁽⁴⁵⁾.

The above mentioned factors, including lack of integration, trust, interpretability, and actionability, limit the use of risk models⁽³⁸⁾ in clinical practice.

As shown in chapter 4 and chapter 6, the model's performance can be lower on data from an external center. Hospitals should have the ability to retrain the model on local datasets if required^(39, 47), to ensure that the model can adapt to local variations in patient demographics, patient outcomes, treatments and other local clinical practices. It is important to note that beyond geographical validation (validation in other institutions: chapters 3, 4, and 6) and domain validation (validation in a new clinical context: chapter 4), temporal validation is required to assess the performance of machine learning models over time⁽⁴⁸⁾. Retraining or updating is needed if the performance of the model decreases over time due to changes in statistical properties of the data (i.e. data drift)⁽⁴⁹⁾. Recently, the first studies have considered elements of the machine learning operations (MLOps) principles. These studies involve deploying a model within the electronic health system, validating it, and monitoring the model over time to determine if retraining is required⁽⁵⁰⁾. In implementation of these models, it is important to look beyond performance statistics (discrimination, calibration), as these can decline due to poor generalization, but also because of lower occurrence of the outcome as a result of improved care⁽⁴⁵⁾. Focus should be on whether outcomes improve, such as decreasing mortality rates among high-risk patients, whether there are changes in who receives treatment and when, and whether the model was responsible for better outcomes. In addition, for effective integration of a risk model in the workflow, there should be a good interplay between the model, available capacity and the action.

To illustrate this, a model may advise (action) that a patient should undergo an ICA procedure due to the high risk of CAD. However, if there is no ICA capacity, the action cannot be performed.



A limited number of machine learning-based risk stratification tools have been implemented for patients with CAD so far. There are multiple barriers to overcome before a model can be implemented into the clinical workflow, including demonstrating clinical efficacy, regulatory compliance, IT integration, and clinical adoption.

Future considerations on AI for risk stratification in CAD

Building upon this thesis, there are several gaps in knowledge that warrant further research.

Future research in CCS

Recent studies have shown that integrating multimodal data with clinical data in a machine learning model provides better prognostic value in predicting outcomes in patients with CAD compared to traditional methods^(15, 51). Further research is needed to explore the full potential of combining clinical, imaging (including TTE data), and genetic data for risk prediction^(15, 51, 52). Additionally, CCS is a progressive disease in which symptoms, risk factors, plaque severity, and treatments for the individual patient may change over time. A time-dependent risk model that takes into account those temporal changes may improve the prediction of further disease progression. This model may have added value compared to existing models that rely on data from a single time point. In **chapter 2 and 3**, we focused on the outcome mortality. Future research on risk stratification should also focus on other major adverse cardiovascular events (MACE) and patient-reported outcomes such as angina⁽⁵³⁾.

Future research in ACS

Machine learning-based risk scores to identify high risk patient with ACS, as presented in **chapter 4**, need further evaluation in untreated patients admitted for (suspected) ACS. Further validation is warranted for patient populations in which the GRACE 2.0 risk score has shown varying discriminative performance, including female patients⁽⁴²⁾ and ethnic minorities⁽⁵⁴⁾. As stated above, it is important to investigate how the use of these models affects clinical decisions and patient outcomes⁽⁵⁵⁾, particularly in comparison to the GRACE 2.0 risk score.

Future research in ICA

The current SYNTAX score uses clinical data and pre-PCI angiographic features to guide clinical decision-making in patients with complex CAD. Training machine learning-based models also on post-PCI angiographic features or images may add value for predicting MACE. This would require a more advanced version of the deep-learning model developed in **Chapter 6**, capable of automatically extracting relevant features from ICA images. Further research may enable (real-time)

lesion-specific recommendations for revascularization, including advice on stent size, length, location, and strategy to standardize procedures and improve PCI outcomes⁽⁵⁶⁾.

Right risk, right time, right place

In context of risk stratification it is important to understand the requirements of the machine learning end-product. In what manner should the model provide the output? To what extent should it be actionable? Should this be in a real-time fashion in the cath lab, in the electronic health record system during a consult, or at the multidisciplinary heart team meeting? Is the outcome risk shown as a percentage or on a dashboard, and is it presented with or without color? The output should be at the right time at the right place of the clinical workflow⁽⁵⁷⁾, as illustrated in Figure 1. To make this all possible, it is essential to have a good collaboration between clinicians, technicians, IT, vendors, patients, and people that can facilitate communication between them, such as technical physicians⁽⁵⁷⁾.

Conclusion

In conclusion, this thesis demonstrates that machine learning offers promising opportunities for risk stratification in patients with CAD, using TTE and ICA data. Machine learning can play an important role in extracting predictive features from these imaging modalities and identifying complex patterns in the data that are predictive for adverse outcomes. Validation of these models in an external center is crucial to determine their generalizability. Further efforts are needed to investigate whether these AI-based risk stratification models can aid clinicians and patients in their decision-making and what impact they have on patient outcomes.

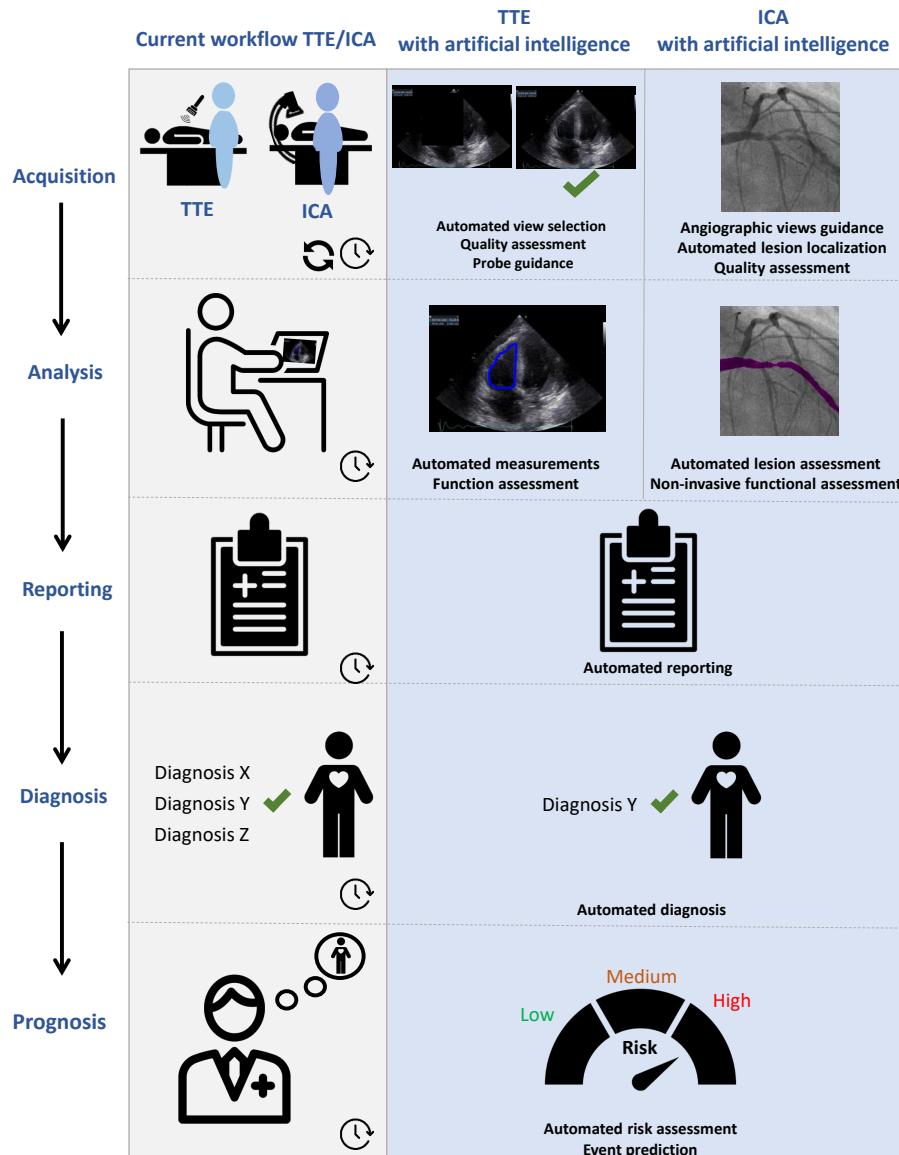


Figure 1: The potential role of artificial intelligence for automated analysis of transthoracic echocardiography (TTE) and invasive coronary angiography (ICA) to support the clinical workflow. Adapted with permission from Schuuring et al., 2021, *Frontiers in Cardiovascular Medicine*⁽¹³⁾.

References

- [1] Vrints C, Andreotti F, Koskinas KC, Rossello X, Adamo M, Ainslie J, et al. 2024 ESC Guidelines for the management of chronic coronary syndromes. *Eur Heart J* 2024;45(36):3415-3537.
- [2] Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, et al. 2023 ESC Guidelines for the management of acute coronary syndromes: Developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J* 2023;44:3720-3826.
- [3] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;41:407-477.
- [4] Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med* 2020;3:10.
- [5] Hughes JW, Yuan N, He B, Ouyang J, Ebinger J, Botting P, et al. Deep learning evaluation of biomarkers from echocardiogram videos. *EBioMedicine* 2021;73:103613.
- [6] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;580:252-256.
- [7] Valsaraj A, Kalmady SV, Sharma V, Frost M, Sun W, Sepehrvand N, et al. Development and validation of echocardiography-based machine-learning models to predict mortality. *eBioMedicine* 2023;90.
- [8] Reddy S. Explainability and artificial intelligence in medicine. *The Lancet Digital Health* 2022;4:e214-e215.
- [9] Kundu S. AI in medicine must be explainable. *Nat Med* 2021;27:1328-1328.
- [10] Yoon CH, Torrance R, Scheinerman N. Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned? *J Med Ethics* 2022;48:581-585.
- [11] Bifari OO. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. *PLoS One* 2023;18:e0284315.
- [12] Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021;113:103655.
- [13] Schuurings MJ, Işgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;8:648877.
- [14] Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38:500-7.
- [15] Pezel T, Sanguineti F, Garot P, Unterseeh T, Champagne S, Toupin S, et al. Machine-Learning Score Using Stress CMR for Death Prediction in Patients With Suspected or Known CAD. *JACC Cardiovasc Imaging* 2022;15:1900-1913.
- [16] Zhang X, Wang X, Xu L, Liu J, Ren P, Wu H. The predictive value of machine learning for mortality risk in patients with acute coronary syndromes: a systematic review and meta-analysis. *Eur J Med Res*. 2023;28(1):451.
- [17] Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol* 2024;42:3-15.
- [18] Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med* 2021;1:1-3.
- [19] Fox KAA, Fitzgerald G, Puymirat E, Huang W, Carruthers K, Simon T, et al. Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ Open* 2014;4:e004425.
- [20] Wenzl FA, Bruno F, Kofoed KF, Raeber L, Roffi M, Stellos K, et al. Validation of the GRACE 3.0 score and redefinition of the risk threshold for early invasive treatment in non-ST-segment elevation acute coronary syndromes: a modelling study from five countries. *European Heart Journal* 2023;44:ehad655.1539.
- [21] Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. *JAMA Network Open* 2022;5:e2233946.
- [22] Grote T, Keeling G. Enabling Fairness in Healthcare Through Machine Learning. *Ethics Inf Technol* 2022;24:39.
- [23] Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11(2):88-94.
- [24] van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. 2021;31(6):3797-3804.
- [25] Du T, Xie L, Zhang H, Liu X, Wang X, Chen D, et al. Training and validation of a deep learning architecture for the automatic analysis of coronary angiography. *EuroIntervention* 2021;17:32-40.
- [26] Neumann FJ, Sousa-Uva M, Ahlsson A, Alfonsi F, Banning AP, Benedetto U, et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J*. 2019;40(2):87-165.
- [27] Popov M, Amanturdieva A, Zhaksylyk N, Alkanov A, Saniyazbekov A, Aimyshev T, et al. Dataset for Automatic Region-based Coronary Artery Disease Diagnostics Using X-Ray Angiography Images. *Sci Data*. 2024;11(1):20.
- [28] Seetharam K, Brito D, Farjo PD, Sengupta PP. The Role of Artificial Intelligence in Cardiovascular Imaging: State of the Art Review. *Front Cardiovasc Med*. 2020;7:618849.
- [29] Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated Learning for Healthcare Informatics. *J Healthc Inform Res*. 2021;5:1-19.
- [30] van Genderen ME, Cecconi M, Jung C. Federated data access and federated learning: improved data sharing, AI model development, and learning in intensive care. *Intensive Care Med*. 2024 Jun;50:974-977.
- [31] Mahendiran T, Thanou D, Senouf O, Meier D, Dayer N, Aminfar F, et al. Deep learning-based prediction of future myocardial infarction

- using invasive coronary angiography: a feasibility study. *Open Heart*. 2023;10(1):e002237.
- [32] Quer G, Arnaout R, Henne M, Arnaout R. Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2021;77(3):300-313.
- [33] Boonstra MJ, Weissenbacher D, Moore JH, Gonzalez-Hernandez G, Asselbergs FW. Artificial intelligence: revolutionizing cardiology with large language models. *Eur Heart J*. 2024;45(5):332-345.
- [34] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) PE/24/2024/REV/1, OJ L, 2024/1689, 12.7.2024.
- [35] van de Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RER. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform*. 2022;29(1):e100495.
- [36] Wenzl FA, Kraler S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *The Lancet* 2022;400:744-756.
- [37] D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *The Lancet* 2021;397:199-207.
- [38] Engel J, van der Wulp I, de Bruijne M, Wagner C. A cross-sectional multicentre study of cardiac risk score use in the management of unstable angina and non-ST-elevation myocardial infarction. *BMJ Open*. 2015;5(11):e008523.
- [39] Aristidou A, Jena R, Topol EJ. Bridging the chasm between AI and clinical implementation. *Lancet*. 2022 Feb 12;399(10325):620.
- [40] Smit IM, Krijthe JH, Kant WMR, Labrecque JA, Komorowski M, Gommers DAMPJ, et al. Causal inference using observational intensive care unit data: a scoping review and recommendations for future practice. *NPJ Digit Med*. 2023;6(1):221.
- [41] Smit JM, Krijthe JH, van Bommel J; Causal Inference for ICU Collaborators. The future of artificial intelligence in intensive care: moving from predictive to actionable AI. *Intensive Care Med*. 2023;49(9):1114-1116.
- [42] Sanger NMR van der, Azzahhafi J, Yin DRPPCP, Peper J, Rayhi S, Walhout RJ, et al. External validation of the GRACE risk score and the risk-treatment paradox in patients with acute coronary syndrome. *Open Heart* 2022;9:e001984.
- [43] Hall M, Bebb OJ, Dondo TB, Yan AT, Goodman SG, Bueno H, et al. Guideline-indicated treatments and diagnostics, GRACE risk score, and survival for non-ST elevation myocardial infarction. *Eur Heart J* 2018;39:3798-3806.
- [44] Saar A, Marandi T, Ainla T, Fischer K, Blöndal M, Eha J. The risk-treatment paradox in non-ST-elevation myocardial infarction patients according to their estimated GRACE risk. *Int J Cardiol* 2018;272:26-32.
- [45] van Amsterdam WAC, van Geloven N, Krijthe JH, Ranganath R, Cinà G. When accurate prediction models yield harmful self-fulfilling prophecies. *Patterns (N Y)*. 2025;6(4):101229.
- [46] Hernán MA, Hsu J, Healy BA, second chance to get causal inference right: a classification of data science tasks. *Chance* 32, 42-49 (2019).
- [47] de Hond AAH, Kant IMJ, Fornasa M, Cinà G, Elbers PWG, Thoral PJ, et al. Predicting Readmission or Death After Discharge From the ICU: External Validation and Retraining of a Machine Learning Model. *Crit Care Med*. 2023;51:291-300.
- [48] de Hond AAH, Shah VB, Kant IMJ, Van Calster B, Steyerberg EW, Hernandez-Boussard T. Perspectives on validation of clinical predictive algorithms. *NPJ Digit Med*. 2023;6(1):86.
- [49] Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol*. 2023;96(1150):20220878.
- [50] Schinkel M, Boerman AW, Paranjape K, Wiersinga WJ, Nanayakkara PWB. Detecting changes in the performance of a clinical machine learning tool over time. *EBioMedicine*. 2023;97:104823.
- [51] Pezel T, Toupin S, Bouisson V, Hamzi K, Hovasse T, et al. A Machine Learning Model Using Cardiac CT and MRI Data Predicts Cardiovascular Events in Obstructive Coronary Artery Disease. *Radiology*. 2025;314(1):e233030.
- [52] Forrest IS, Petrazzini BO, Duffy Á, Park JK, Marquez-Luna C, Jordan DM, et al. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. *Lancet*. 2023;401(10372):215-225.
- [53] Collison D, Copt S, Mizukami T, Collet C, McLaren R, et al. Angina After Percutaneous Coronary Intervention: Patient and Procedural Predictors. *Circ Cardiovasc Interv*. 2023;16(4):e012511.
- [54] Moledina SM, Kontopantelis E, Wijeyesundara HC, Banerjee S, Van Spall HGC, Gale CP, et al. Ethnicity-dependent performance of the Global Registry of Acute Coronary Events risk score for prediction of non-ST-segment elevation myocardial infarction in-hospital mortality: nationwide cohort study. *Eur Heart J* 2022;43:2289-2299.
- [55] Gale CP, Stocken DD, Aktaa S, Reynolds C, Gilberts R, Brieger D, et al. Effectiveness of GRACE risk score in patients admitted to hospital with non-ST elevation acute coronary syndrome (UKGRIS): parallel group cluster randomised controlled trial. *BMJ*. 2023;381:e073843.
- [56] van Beek KAJ, Timmermans MJC, Derkx L, Cheng JM, Kraaijeveld AO, Arkenbout EK, et al. Contemporary Use of Post-Dilatation for Stent Optimization During Percutaneous Coronary Intervention; Results From the Netherlands Heart Registration. *Catheter Cardiovasc Interv*. 2025;105(4):870-877.
- [57] Kim B, Romeijn S, van Buchem M, Mehrizi MHR, Grootjans W. A holistic approach to implementing artificial intelligence in radiology. *Insights Imaging*. 2024;15(1):22.



[Summary](#)
[Samenvatting \(Dutch summary\)](#)
[List of publications](#)
[Portfolio](#)
[Acknowledgments](#)
[Curriculum Vitae/About the author](#)

Summary

Coronary artery disease affects millions of individuals worldwide and has high mortality rates, despite advanced medical and interventional treatment. The stratification of patients based on the likelihood of experiencing an adverse outcome, also called risk stratification, is the cornerstone for treatment decision in patients with coronary artery disease. It assists clinicians in identifying high-risk or rising-risk patients and facilitates resource allocation by selecting those who may benefit most from more intensive care.

Machine learning models, models that learn from data to make predictions, may improve the risk stratification of patients with coronary artery disease. Machine learning is a subfield of artificial intelligence, which refers to the broad concept of computer systems performing tasks that previously required human intelligence. In this thesis, we have investigated the value of machine learning for risk stratification of patients with coronary artery disease, with specific focus on the imaging modalities transthoracic echocardiography and invasive coronary angiography data.

Current chronic coronary syndrome guidelines recommend transthoracic echocardiography in all patients with chronic coronary syndrome (stable patients with coronary artery disease) for the measurement of the left ventricular function, which is considered to be one of the strongest determinants of mortality. Transthoracic echocardiography encompasses more information than the left ventricular function alone. For example, transthoracic echocardiography is the key diagnostic tool for assessment of valvular structure and flow, which is essential for the detection of valvular heart disease. In **chapter 2**, the impact of valvular heart disease on mortality was investigated in 1984 patients with chronic coronary syndrome in a retrospective cohort study. The results showed that a total of 325 patients with chronic coronary syndrome had moderate valvular heart disease, while 44 patients had severe valvular heart disease. Moderate or severe valvular heart disease was associated with an increased risk of mortality, independent of left ventricular function and other risk factors. After multivariable correction, moderate tricuspid regurgitation was independently associated with mortality in patients with chronic coronary syndrome. These findings demonstrate the importance of echocardiographic assessment of valvular heart disease for risk stratification, in addition to left ventricular function, in patients with chronic coronary syndrome.

Machine learning models trained on both clinical and transthoracic echocardiography data may improve the risk prediction of patients with chronic coronary syndrome. This hypothesis was strengthened by studies on machine learning-based risk scores using data of other imaging modalities (coronary computed tomographic, stress cardiac magnetic resonance). These studies showed that machine learning models better discriminated between patients who are likely to die and those who are not. In **chapter 3**, we present the first study on the accuracy of machine learning using clinical and transthoracic echocardiography data to predict five-year mortality in patients with chronic coronary syndrome. The model, an Extreme Gradient Boosting model, was trained on a retrospective cohort of 775 patients with chronic coronary syndrome and demon-

strated a good performance, superior to traditional risk stratification tools. The machine learning model demonstrated good performance on patients from a different tertiary center. Left ventricular dysfunction and significant tricuspid regurgitation were included as predictors in the machine learning model. These findings suggest that machine learning may support clinicians in assessing the risk of mortality in patients with chronic coronary syndrome.

In **chapter 4**, we validated two machine learning-based risk scores for predicting mortality, the GRACE 3.0 risk score and the PRAISE risk score, in patients with acute coronary syndrome treated with percutaneous coronary intervention. In the 2759 patients with non-ST-elevation acute coronary syndrome treated with percutaneous coronary intervention, the GRACE 3.0 risk score was more effective in predicting in-hospital mortality compared to the GRACE 2.0 risk score currently recommended by guidelines. The PRAISE risk score, validated in 4347 patients with acute coronary syndrome treated with percutaneous coronary intervention, demonstrated a lower performance compared to the study that developed the model. The PRAISE risk score did not improve the prediction of one-year mortality compared to the GRACE 2.0 risk score. These findings provide suggestive evidence about the performance of these models in an external center.

Invasive coronary angiography is recommended in patients with a high risk of cardiovascular events or patients with severe coronary artery disease. Typically, invasive coronary angiography images are visually assessed, which is prone to subjective interpretation. Machine learning may assist clinicians in interpreting these images. In **chapter 5**, we describe the current state and future perspectives of machine learning for automated imaging analysis in invasive coronary angiography. In the twelve studies reviewed, machine learning solutions were reported for automated frame selection, segmentation, lesion detection and assessment, and functional assessment of coronary flow. Only three out of the twelve machine learning models were validated on data of an external center. To evaluate the performance of invasive coronary angiography interpretation models across various clinical settings, the aim of **chapter 6** was to train and validate advanced models (deep learning models). These models were trained to segment coronary arteries and detect significant stenoses on invasive coronary angiography images, and were validated in an external center. Deep learning-based models were created using data from two tertiary centers, and both quantitative and qualitative analyses were performed. A total of 10,573 invasive coronary angiography images were used to train models: 9065 from Center 1 (2624 patients) and 1508 (456 patients) from Center 2. The models achieved performance levels on par with experts in the segmentation of coronary arteries and detection of significant stenoses in the main arteries. In addition, we found important factors that need to be taken into account when developing models for automated coronary angiography analysis, including dataset size and prevalence of stenoses in the data. This study laid the foundation for further research on automated analysis of invasive coronary angiography images. It enables the quantification of features related to the severity of coronary artery disease, which are crucial for assessing the risk of cardiovascular events. In addition, it provides opportunities for treatment planning and interventional guidance (**chapter 5**).

Chapter 7 outlines a general discussion with answers on the main research questions, reflec-

tions on the findings, and recommendations for future work. Machine learning, and artificial intelligence in general, offer opportunities to improve risk stratification in patients with coronary artery disease. Machine learning can play an important role in extracting predictive features from imaging modalities (transthoracic echocardiography and invasive coronary angiography), and identifying complex patterns in the data that are predictive for adverse outcomes. Validation of these models in an external center is crucial to determine their generalizability. Further efforts are needed to investigate whether these artificial intelligence-based risk stratification models can aid clinicians and patients in their decision-making and what impact these models have on patient outcomes.

Samenvatting (Dutch summary)

Coronaire hartziekte, ook wel kransslagadervernauwing genoemd, is een veelvoorkomende aandoening die miljoenen mensen wereldwijd treft en vaak dodelijk kan zijn, ondanks geavanceerde behandelingen. Om te beslissen welke behandeling het beste is voor patiënten met deze ziekte, gebruiken artsen een methode genaamd risicostratificatie. Dit betekent dat ze patiënten indelen op basis van hun risico op een slechte uitkomst. Hierdoor kunnen artsen beter bepalen welke patiënten intensievere zorg nodig hebben en het meeste baat zullen hebben bij bepaalde behandelingen.

Machine learning-modellen, die leren van gegevens om voorspellingen te doen, kunnen de risicostratificatie van patiënten met coronaire hartziekten mogelijk verbeteren. Machine learning is een deelgebied van kunstmatige intelligentie, dat verwijst naar het brede concept van computersystemen die taken uitvoeren die voorheen menselijke intelligentie vereisten. In dit proefschrift hebben we de waarde van machine learning voor de risicostratificatie van patiënten met coronaire hartziekte onderzocht, met specifieke focus op de beeldvormingstechnieken transthoracale echocardiografie en invasieve coronaire angiografie-gegevens.

De huidige richtlijnen voor chronisch coronair syndroom (stabiele patiënten met kransslagadervernauwing) bevelen transthoracale echocardiografie aan bij alle patiënten met chronisch coronair syndroom voor het meten van de linkerventrikelfunctie, die wordt beschouwd als een van de sterkste determinanten van mortaliteit. Transthoracale echocardiografie bevat meer informatie dan alleen de linkerventrikelfunctie. Transthoracale echocardiografie is bijvoorbeeld het belangrijkste diagnostische hulpmiddel voor de beoordeling van de klepstructuur en bloedstroom door de kleppen, wat essentieel is voor de detectie van hartklepaandoeningen. In **hoofdstuk 2** werd de impact van hartklepaandoeningen op mortaliteit onderzocht bij 1984 patiënten met chronisch coronair syndroom in een retrospectieve cohortstudie. De resultaten toonden aan dat in totaal 325 patiënten met chronisch coronair syndroom matige hartklepaandoeningen hadden, terwijl 44 patiënten ernstige hartklepaandoeningen hadden. Matige of ernstige hartklepaandoeningen werden geassocieerd met een verhoogd risico op mortaliteit, onafhankelijk van de linkerventrikelfunctie en andere risicofactoren. Na het uitvoeren van multivariabele correcties werd matige tricuspidalregurgitatie geassocieerd met een verhoogd risico op mortaliteit bij patiënten met chronisch coronair syndroom. Deze bevindingen tonen het belang aan van echocardiografische beoordeling van hartklepaandoeningen voor risicostratificatie bij patiënten met chronisch coronair syndroom, naast de linkerventrikelfunctie.

Machine learning-modellen die zijn getraind op zowel klinische als transthoracale echocardiografie-gegevens, kunnen de risicovoorspelling van patiënten met chronisch coronair syndroom mogelijk verbeteren. Deze hypothese werd versterkt door studies naar op machine learning gebaseerde risicoscores, ontwikkeld met gegevens van andere beeldvormende modaliteiten (coronaire CT-angiografie, MRI-stressonderzoek). Deze studies toonden aan dat machine learning-modellen beter onderscheid maakten tussen patiënten die waarschijnlijk zouden sterven en patiënten die dat niet zouden doen. **Hoofdstuk 3** beschrijft de eerste studie

naar de nauwkeurigheid van machine learning met behulp van klinische en transthoracale echocardiografie-gegevens om de mortaliteit na vijf jaar bij patiënten met chronisch coronair syndroom te voorspellen. Het model, een Extreme Gradient Boosting-model, werd getraind op een retrospectief cohort van 775 patiënten met chronisch coronair syndroom en toonde een goede prestatie, beter dan de gebruikelijke risicostratificatiertools. Het machine learning-model leverde ook goede prestaties bij patiënten in een ander academisch ziekenhuis. Linkerventrikeldysfunctie en tricuspidalisklepinsufficiëntie, een aandoening van de tricuspidalisklep, werden opgenomen als voorspellers in het machine learning-model. Deze bevindingen wekken de suggestie dat machine learning clinici kan ondersteunen bij het beoordelen van het risico op sterfte bij patiënten met chronisch coronair syndroom.

In hoofdstuk 4 valideerden we twee op machine learning gebaseerde risicoscores voor het voorspellen van overlijden, de GRACE 3.0 en de PRAISE-risicoscore bij patiënten met acuut coronair syndroom (onstabiele patiënten met kransslagadervernauwing) die werden behandeld met een dotterbehandeling. Bij de 2759 patiënten met acuut coronair syndroom zonder ST-elevatie die werden behandeld met de dotterbehandeling was de GRACE 3.0 risicoscore een effectievere risicovoorspellingstool vergeleken met de oudere GRACE 2.0-score voor het voorspellen van overlijden in het ziekenhuis. De PRAISE risicoscore werd gevalideerd bij 4347 patiënten met acuut coronair syndroom die werden behandeld met een dotterbehandeling en toonde een lagere prestatie vergeleken met de studie die het model ontwikkelde. De PRAISE score verbeterde de voorspelling van overlijden na één jaar niet vergeleken met de GRACE-2.0 risicoscore. Deze bevindingen leveren indicatief bewijs voor de prestaties van deze modellen in een ander ziekenhuis.

Invasieve coronaire angiografie wordt aanbevolen bij patiënten met een hoog risico op cardiovasculaire gebeurtenissen of patiënten met ernstige coronaire hartziekte. Meestal worden de invasieve coronaire angiografie-beelden visueel beoordeeld, wat vatbaar is voor verschil in interpretatie tussen beoordelaars. Machine learning kan clinici helpen bij het interpreteren van deze beelden. In hoofdstuk 5 beschrijven we de huidige staat en toekomstige perspectieven van machine learning voor geautomatiseerde beeldanalyse in invasieve coronaire angiografie. In de twaalf beoordeelde onderzoeken werden machine learning-oplossingen beschreven gefocust op geautomatiseerde beeldselectie, segmentatie, detectie en beoordeling van vernauwingen, en functionele beoordeling van de coronaire bloeddoorstroming. Slechts drie van de twaalf machine learning-modellen werden gevalideerd op gegevens van een ander ziekenhuis. Om te beoordelen in hoeverre invasieve coronaire angiografie-interpretatiemodellen generaliseerbaar zijn naar verschillende klinische settings, was het doel van hoofdstuk 6 om geavanceerde modellen (deep learning-modellen) te trainen en valideren. Deze modellen werden getraind om kransslagaders te segmenteren en significante vernauwingen op invasieve coronaire angiografie beelden te detecteren. Deze modellen werden gemaakt met behulp van gegevens van twee academische ziekenhuizen, en zowel kwantitatieve als kwalitatieve analyses werden uitgevoerd. In totaal werden 10573 invasieve coronaire angiografie-afbeeldingen gebruikt om modellen te trainen: 9065 van ziekenhuis 1 (2624 patiënten) en 1508 (456 patiënten) van ziekenhuis 2. De prestaties van de modellen waren vergelijkbaar met die van experts in de segmentatie van kransslagaders

en detectie van vernauwingen in de hoofdslaagaders. Daarnaast vonden we belangrijke factoren waarmee rekening moet worden gehouden bij het ontwikkelen van modellen voor automatische analyse van coronaire angiografie beelden, waaronder de grootte van de dataset en hoe vaak vernauwingen voorkomen in de data. Deze studie heeft een basis gelegd voor verder onderzoek naar het automatisch analyseren van beelden van invasieve coronaire angiografie om vernauwingen kwantitatief te beoordelen. Het maakt het mogelijk om kenmerken te kwantificeren die verband houden met de ernst van coronaire hartziekte, wat cruciaal is voor het inschatten van het risico op cardiovasculaire gebeurtenissen. Daarnaast biedt het mogelijkheden voor behandelplanning en interventieondersteuning (**hoofdstuk 5**).

Hoofdstuk 7 biedt een algemene discussie met antwoorden op de belangrijkste onderzoeks vragen, reflecties op deze bevindingen, en aanbevelingen voor toekomstig onderzoek. Machine learning, en kunstmatige intelligentie in het algemeen, bieden kansen om de risicostratificatie bij patiënten met coronaire hartziekte te verbeteren. Machine learning kan een belangrijke rol spelen bij het extraheren van voorspellende kenmerken uit beeldvormende technieken (transthoracale echocardiografie en invasieve coronaire angiografie) en bij het herkennen van complexe patronen in de gegevens die voorspellend zijn voor slechte uitkomsten. Validatie van deze modellen in een ander ziekenhuis is cruciaal om hun generaliseerbaarheid te bepalen. Verdere onderzoeken zijn nodig om te bepalen of deze op kunstmatige intelligentie gebaseerde risicostratificatiemodellen clinici en patiënten kunnen helpen bij hun besluitvorming, en welke impact ze hebben op patiëntuitkomsten.

List of publications

Publications in this thesis

Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, et al. Current State and Future Perspectives of Artificial Intelligence for Automated Coronary Angiography Imaging Analysis in Patients with Ischemic Heart Disease. *Curr Cardiol Rep* 2022;24:365–376.

Molenaar MA, Bouma BJ, Coerkamp CF, Man JP, Išgum I, Verouden NJ, et al. The impact of valvular heart disease in patients with chronic coronary syndrome. *Front Cardiovasc Med* 2023;10:1211322.

Molenaar MA, Bouma BJ, Asselbergs FW, Verouden NJ, Selder JL, Chamuleau SAJ, et al. Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome. *Eur Heart J Digit Health*. 2024;5:170-182.

Molenaar MA, Selder JL, Schmidt AF, Asselbergs FW, Nieuwendijk JD, van Dalfsen B, et al. Validation of machine learning-based risk stratification scores for patients with acute coronary syndrome treated with percutaneous coronary intervention. *Eur Heart J Digit Health*. 2024;5:702-711.

Molenaar MA, Hebb E, Selder JL, Shekiladze N, Sandesara PB, Nicholson WJ et al. Deep Learning-Based Segmentation of Coronary Arteries and Stenosis Detection in X-ray Coronary Angiography. Submitted to *JACC: Advances*.

Other Publications

van der Meijden S, **Molenaar MA**, Somhorst P, Schoe A. Calculating mechanical power for pressure-controlled ventilation. *Intensive Care Med.* 2019;45(10):1495-1497.

Grävare Silbernagel K, Malliaras P, de Vos RJ, Hanlon S, **Molenaar MA**, Alfredson H, et al. ICON 2020-International Scientific Tendinopathy Symposium Consensus: A Systematic Review of Outcome Measures Reported in Clinical Trials of Achilles Tendinopathy. *Sports Med.* 2022;52(3):613-641.

Molenaar MA, Coerkamp CF, Man JP, Bouma BJ, Isgum I, Verouden NJ, et al. The impact of valvular heart disease in patients referred for suspected coronary artery disease. *European Heart Journal - Cardiovascular Imaging.* 2022;23(Supplement_1):jeab289.234.

Molenaar MA, Bouma BJ, Isgum I, Verouden NJ, Selder JL, Chamuleau SAJ, et al. Predictive value of the left ventricular function in coronary artery disease: should we tailor risk-stratification for men and women? *European Heart Journal.* 2021;42(Supplement_1):ehab724.1152.

Chien W, Rodriguez Rivero C, Haas SD, **Molenaar MA**. Echocardiographic Clustering by Machine Learning in Children with Early Surgically Corrected Congenital Heart Disease. ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH). 2023.

Vink CEM, de Jong EAM, Woudstra J, **Molenaar MA**, Kamp O, Götte MJW, et al. The role of myocardial blood volume in the pathophysiology of angina with non-obstructed coronary arteries: The MICORDIS study. *Int J Cardiol.* 2024;415:13247

Jansen GE, de Vos BD, **Molenaar MA**, Schuuring MJ, Bouma BJ, et al. Automated echocardiography view classification and quality assessment with recognition of unknown views. *J Med Imaging (Bellingham).* 2024;11:054002.

Jansen GE, **Molenaar MA**, Schuuring MJ, Bouma BJ, Isgum I. Automated mitral valve segmentation in PLAX-view transthoracic echocardiography for anatomical assessment and risk stratification. *Comput Biol Med.* 2025 Sep;196(Pt C):110900.

Portfolio

PhD candidate: Mitchel Molenaar

Period: 2020 to 2025

Supervisors: Prof. dr. S.A.J. Chamuleau

Co-supervisors: Dr. M.J. Schuuring

Dr. N.J. Verouden

PhD training	Year	ECTS
Courses		
eBROK	2020	1.0
Research Data Management	2020	0.9
Scientific Writing in English for Publication	2020	1.5
Peer to Peer Group coaching	2021	0.5
Project management	2021	1.3
Computing in R	2021	1.4
Practical Biostatistics	2021	2.1
Advanced Topics in Biostatistics	2022	1.5
Personal development course (SUAS)	2023	1.5
Assertivity course (SUAS)	2023	1.5
Seminars, Workshops, and Master Classes		
Meetup: Connecting AI and Medical Imaging	2020	0.1
MICCAI: Workshop	2023	0.3
Rotterdam E-health ontbijt	2023	0.1
Wetenschapsavond: AI in BEELD	2024	0.1
(International) Conferences and Presentations		
ESC Digital Health Week (online)	2020	0.5
NVVC congress spring (presentation online)	2021	0.7
EuroEcho (presentation online)	2021	1.5
ESC congress (poster online)	2021	0.7
Morning meeting (presentation)	2021	0.7
Technical innovations in medicine congress (online)	2021	0.3
NVVC congress spring (presentation)	2022	1
NVVC congress fall (presentation)	2022	1
ACS symposium (presentation)	2022	1
ACV meeting (presentation)	2022	0.7
ICT&Health (presentation)	2022	1
Morning meeting (presentation)	2023	0.7
EuroPCR, Paris (poster)	2023	1.5
7th DCVA-NLHI Translational Cardiovascular Research meeting (presentation)	2023	1
IFL Symposium Healthy AI, Heidelberg	2023	0.5
MICCAI, Vancouver (presentation)	2023	1.5
NVVC-congress fall (presentation)	2023	1
ACS symposium (presentation)	2023	1

continued

PhD training	Year	ECTS
Tutoring, Mentoring		
Casper Coerkamp, Master student Medicine	2021	1.5
Anass Mokhtari, Master student Medicine	2021	1.5
Will Chien, Master Data Science	2022	1.5
Hafsa Rahimbaksh, Master student Medicine	2022	1.5
Jelle Nieuwendijk, Master student Medicine	2022	1.5
Brigitte van Dalsen, Master student Medicine	2023	1.5
Saskia Dahmen, Master student Medicine	2023	1.5
Supervising		
Harry Gibson, CATHAI project	2021-2023	0.2
Verne Ros, CATHAI project	2021	0.2
Roy van Erck, CATHAI project	2021	0.2
Chaimae Bouchnaf, CATHAI project	2021-2022	0.2
Hilal Sarica, CATHAI project	2021-2022	0.2
Afifa Amjad, CATHAI project	2021-2022	0.2
Jelle Nieuwendijk, CATHAI project	2021-2024	0.2
Kenneth Schweikart, CATHAI project	2021	0.2
Bodessa Overbeek, CATHAI project	2021-2022	0.2
Anouk Ree, CATHAI project	2021-2024	0.2
Jessica Lijsterink, CATHAI project	2021-2023	0.2
Ji-rong Kruijswijk, CATHAI project	2021-2022	0.2
Anna-Sophia ten Voorde, CATHAI project	2021-2022	0.2
Rowan Puijk, CATHAI project	2021-2024	0.2
Charlotte Garstman, CATHAI project	2022-2024	0.2
Justin Souprayan, CATHAI project	2022-2024	0.2
Claire Veldkamp, CATHAI project	2022-2024	0.2
Denise van Domburg, CATHAI project	2023-2024	0.2
Pieter Schoots, CATHAI project	2022	0.2
Thomas van Scholkema, CATHAI project	2022	0.2
Nadine Geldof, CATHAI project	2022	0.2
Storm van Paesen, CATHAI project	2023	0.2
Mei-Juan Brand, CATHAI project	2023	0.2
Rahul Bhoera, CATHAI project	2023	0.2
Ariana Mihaita, CATHAI project	2023	0.2
Sara Jalaly, CATHAI project	2023	0.2
Merel Bekkers, CATHAI project	2023	0.2
Kevin Heideman, CATHAI project	2023	0.2
Grants, Awards, and Prizes		
UvA Travel Grant	2023	
CLINICCAI Travel Grant	2023	
Third prize 7th DCVA-NLHI Research meeting	2023	

Acknowledgments

“Je zult met veel mensen moeten samenwerken. Kun je goed samenwerken?” Dat was een van de eerste vragen tijdens mijn sollicitatiegesprek voor de promotieplek. Wat als ik toen ‘nee’ had geantwoord? Dan was dit boekje waarschijnlijk een stuk korter geweest, en dit dankwoord al helemaal. Zonder de inzet en hulp van velen was dit proefschrift nooit tot stand gekomen. Daar ga ik iedereen eens goed voor bedanken.

Beste Steven, ik wil je bedanken voor de interesse die je vanaf het begin in mij hebt getoond. Je bent een echt mensenmens en was altijd geïnteresseerd in hoe het met mij ging. Ik vergeet nooit de woorden: “Mitch, als er wat is, kun jij mij altijd appen hè.” Hoewel we elkaar niet wekelijks zagen, nam je echt de tijd voor me tijdens onze afspraken. Daarnaast kon je kritisch zijn, vooral wanneer zaken niet duidelijk waren, wat dit proefschrift zeer ten goede is gekomen. Dank voor het vertrouwen dat je in mij hebt gegeven en de nodige extra maanden die je mogelijk hebt gemaakt.

Beste Niels, wat ben jij een fijne vent in de omgang. Het vertrouwen dat je vanaf het begin in mij had, waardeer ik enorm. Jouw inzet om mij bij Philips over de vloer te krijgen en te gaan ‘kruisbestuiven’ heeft mij als technisch geneeskundige enorm laten groeien. Ik keek altijd uit naar het CATHAI-uurtje samen met Jasper op donderdagmiddag, waarbij er ook voldoende ruimte was voor humor en plezier. Hoewel ik niet al je vele ideeën kon oppakken, heeft dit me altijd gestimuleerd om samen te werken met anderen en na te denken over de mogelijke toepassingen van AI in het cathlab. Dank voor al je steun en vertrouwen de afgelopen jaren!

Beste Mark, ik ken maar weinig mensen die zo hard werken als jij. Hoe laat ik ook een mailtje naar je stuurde, ik kon erop vertrouwen dat ik de volgende ochtend een reactie had. Ik wil je bedanken voor hoe je mij hebt bijgestaan. Ik kon ieder moment van de week laagdrempelig bij je langskomen, en je wilde echt dat onze projecten een succes werden. Je hands-on mentaliteit is echt complimentwaardig: “zal ik een gesprek plannen?” “Welnee, we bellen degene nu gewoon in”. Ook je opmerking nadat we een moeizaam gesprek hadden, “zonder wrijving geen glans”, heeft me gemotiveerd om door te zetten.

Beste Jasper, ik wil je graag bedanken voor je betrokkenheid. Je stelde vaak goede vragen en wilde altijd begrijpen hoe iets werkt, wat mij heeft geholpen om dingen helder te krijgen. Daarnaast waardeer ik je optimisme, dat op een zeer positieve manier heeft bijgedragen aan het CATHAI-project!

Beste Berto, ik wil je enorm bedanken voor de interesse die je in mij hebt getoond. Ik kon altijd bij je terecht met vragen, hoe druk je het ook had. Je bent direct en herinnerde me eraan om progressie te blijven boeken. Daarnaast weet je altijd snel de vinger op de zere plek te leggen, wat ervoor heeft gezorgd dat mijn manuscripten steeds beter werden.

Beste Folkert, gedurende mijn promotie heb ik je leren kennen en zijn we steeds meer naar elkaar toegegroeid. Ik waardeer dat ook jij de toegevoegde waarde van de technisch ge-

neeskundige ziet. Bedankt voor je hulp bij het aanvragen van beurzen, het geven van feedback op presentaties en het verbeteren van mijn artikelen!

Dear Ivana, even though our paths diverged, I would like to thank you for your support during the first phase of my PhD and for your contribution to the second chapter.

Beste leden van de promotiecommissie, beste Prof. dr. A. Abu-Hanna, Prof. dr. M.P. Schijven, Prof. dr. F.M.A.C. Martens, Prof. dr. R. Nijveldt, Dr. J.M. Wolterink, en Prof. dr. H.A. Marquer- ing, hartelijk dank voor het beoordelen van mijn proefschrift en voor jullie bereidheid om deel te nemen aan de promotiecommissie.

Beste Philips-team, beste Javier, Martijn, Britt, Vincent, Yida, Alex, Neil, Iris, Jelle, en alle anderen met wie ik heb samengewerkt. Ik wil jullie bedanken voor de fijne samenwerking in de afgelopen jaren. Ik ben enorm dankbaar voor de ‘kruisbestuiving’ en ben trots dat we CATHAI samen zo’n succes hebben kunnen maken.

Beste Sylvia, Anita, Regina, en Lydia, hartelijk dank voor jullie ondersteuning bij de adminis- tratie, de financiële afhandeling, het plannen van vergaderingen en het boeken van vergader- ruimtes. Jullie inzet heeft het werk voor mij een stuk makkelijker en plezieriger gemaakt.

Beste Jelle, als kamergenoten met dezelfde technisch-geneeskundige achtergrond konden wij goed met elkaar opschieten. Ik wil je bedanken voor de waardevolle discussies, het delen van elkaars uitdagingen en de gezellige momenten die we samen hebben gehad!

Beste Gino, in het begin van onze samenwerking had ik het gevoel dat er een bepaalde politieke muur tussen ons stond, die we langzaam hebben afgebroken. Ik wil je bedanken voor de fijne samenwerking en de koffiemomentjes die we zo nu en dan deelden. Het is echt complimentwaardig wat je allemaal hebt weten te bereiken met de echobeelden.

Beste mede-PhD’ers, ook al was ik er niet altijd door mijn gependel tussen AMC, VUMC en Philips, wil ik jullie bedanken voor de gezellige momenten de afgelopen jaren!

Beste Machteld, Floriaan, Philip, Stefan, Stephan, Marijke, Sean, Clara, Larissa, Vito, Alicia, Joyce en Marion, we hebben elkaar wat later tijdens mijn promotieonderzoek leren kennen, maar ik wil jullie bedanken voor de gezelligheid en de inhoudelijke vragen die ik bij jullie kon stellen. Het is geweldig om te zien hoe data steeds belangrijker wordt in het hartcentrum en hoe hard iedereen eraan werkt.

Beste Roderick, ik ben blij dat we na het afronden van mijn studie de banden weer hebben kunnen aanhalen. Jouw praktische manier van denken is voor mij een groot voorbeeld geweest. Ik hoop in de toekomst contact met je te blijven houden en samen te kunnen werken!

Dear Elsa, Pratik, Syed, and the rest of the Emory team, I want to sincerely express my grati- tude for our collaboration. It is thanks to your contributions that we were able to achieve such an impactful publication.

Beste Geneeskunde studenten, beste Jelle, Brigitte, Casper, Anass, Hafsah en Saskia, bedankt

voor jullie inzet. Hoewel het opschonen en aanvullen van de databases een tijdrovende en soms vermoedende klus was, hebben jullie allemaal fantastisch werk geleverd. Dankzij jullie begeleiding heb ik zelf ook veel kunnen leren.

Beste CATHAI-studenten, beste Jelle, Charlotte, Denise, Claire, Anouk, Mei-Juan, en de andere 23 studenten, zonder jullie was het nooit gelukt om alle coronairangiogrammen van 3.000 patiënten in te tekenen. Petje af voor jullie ijver en nauwkeurigheid!

Beste Merijn, Ernest, Tom, bedankt voor jullie hulp bij de uitgifte van de echodata. Zonder jullie hulp zou dit niet gelukt zijn. Beste Paul, Remko, Bas, Maryam, Christiaan, Daniël, Nico en Ralf, bedankt voor jullie hulp bij de uitgifte van de patiëntdata.

Beste Matthijs, Björn, Paulo, Brian, Fleur, Jules en Viktor, bedankt voor het delen van jullie ervaringen bij het opschalen van AI op de verschillende afdelingen. Ik heb veel geleerd van ieders aanpak en inzichten.

Beste Raoul, Rob, Iris, Nazma, en Maaike, bedankt voor jullie juridische ondersteuning.

Verder wil ik alle co-auteurs bedanken voor hun waardevolle bijdrage aan de artikelen in dit proefschrift. Dankzij jullie inzet is de kwaliteit van de artikelen aanzienlijk verbeterd.

Ook mijn Thuisarts-collega's wil ik niet vergeten. Beste Thuisarts-collega's, bedankt voor de fijne samenwerking die we hebben, en jullie interesse in de voortgang van mijn boekje.

Mijn lieve vrienden wil ik bedanken voor de nodige afleiding die ze mij hebben geboden. Lieve Timo, Gaston, Tim, Ruben, Denzel, Jeroen, Ingmar en de lieve aanhang daarvan, ik ben enorm dankbaar voor onze vriendschap, en voor hoe jullie mij waarderen om wie ik ben. Ons samenzijn geeft me veel energie en zorgt altijd voor plezier en gelach. Ik hoop nog vele jaren met jullie op pad te gaan, de podgarsten op te nemen, jullie te voorzien van heerlijke gratis drankjes achter de festivalbar, en samen hardlooptochtjes te maken. Alf Leila Wa Leila! Beste Wandeljannen, Milan, Thijs, Jim, ik heb enorm genoten van onze studiereüniertjes. Hoewel je aan het eind van de avond niet super veel weet van de ander, weet je wel zeker dat je gelachen hebt.

Lieve Jeffrey, we zien elkaar niet vaak, maar als we elkaar zien is het altijd feest. Ik ga graag snel weer met je lopen.

Lieve Timo en Ingmar, bedankt dat jullie op mijn promotiedag willen nymphomanen.

Lieve Lilian en Jesse, Waddinxveen is misschien niet prachtig, maar dankzij jullie voelt het als een geweldige plek. Bedankt voor de oprechte interesse die jullie hebben getoond in mijn PhD. Ik kijk nu alweer uit naar onze volgende hardloopsessie, het spikeballen met de rest van de 'fam', en natuurlijk jullie volgende bruiloft.

Lieve Anouk en Daan, ik vind het echt superleuk om met jullie af te spreken. Fijn dat jullie steeds een beetje dichterbij komen wonen!

Lieve buurtjes, bedankt voor de leuke nieuwe contacten en de warme ontvangst in de buurt. Lieve Bob, dankjewel voor het sparren en je input tijdens de afronding van dit boekje.

Lieve Remi en familie, wie had ooit kunnen dromen dat die decemberdag in de trein het begin zou zijn van zo'n bijzonder verhaal? Nu, maanden later, kan ik je dit boekje overhandigen en ben je bij mijn promotie. Wat een cadeau is dat! Het meisje op de achterkant symboliseert niet alleen het nieuwe leven dat voor je ligt, maar herinnert mij er elke dag aan om met de verwondering van een kind door het leven te gaan. Ik wil je enorm bedanken voor je bijdrage aan dit boekje!

Lieve bonusfamilie, wat ben ik dankbaar dat ik deel mag uitmaken van jullie familie. Lieve Leo en Anita, ik denk dat er maar weinig huizen zijn waar ik me zó thuis voel. Ik kan bij wijze van spreken bij jullie op tafel schijten, wat jullie nog goed zouden vinden ook. De komende maanden zullen pittig worden. Weet dat we er altijd voor jullie zijn! Lieve Tanja, ik ben blij dat ondanks je bemoedigende woorden "Hoe is het toch mogelijk dat jij promoveert, clown?" en "je bent niet goed wijs", het toch gelukt is. Lieve Axelle en Pieter, ik ben enorm dankbaar dat we afgelopen jaren zo naar elkaar toe zijn gegroeid. Komen jullie straks bij ons oppassen? Lieve Lyonne en Joël, jullie zijn voor mij echt een voorbeeld hoe je van het leven geniet en de dingen doet die je écht leuk vindt. Bedankt, Lyonne, voor het ontwerpen van de mooie kaft van dit boekje! Lieve Richard, bedankt dat we in drukke tijden konden genieten van je kookkunsten en vaak zo konden aanschuiven. Lieve Oma van Gaalen, bedankt dat ik dit boekje grotendeels aan de Spanjessingel heb mogen schrijven.

Superlatieven schieten tekort als ik over mijn familie praat. Lieve Joyce en Simon, na regen komt zonneschijn, en die zonneschijn gaat er voor jullie zéker komen. Jullie zijn prachtige, lieve mensen, en ik schuif met veel plezier aan op de donderdagavond. Lieve Melanie en Erwin, jullie maken mij altijd aan het lachen, en daar geniet ik echt heel erg van. Als Joyce en Simon mij niet willen hebben, kom ik graag bij jullie mee-eeten. Lieve oma, Lieve Nol, ik ben dankbaar dat ik u dit boekje mag overhandigen. Opa zou hier vast heel trots op zijn. Lieve papa en mama, vroeger zeiden jullie vaak: "Wat het eindresultaat ook is, als we maar zien dat je je best doet, zijn we enorm trots." Tegenwoordig denk ik dat: "Wat je ook doet, we zijn enorm trots," nog beter bij jullie past. Voor dit boekje zijn jullie geërfde eigenschappen mij zeer ten goede gekomen: de onrust, het altijd bezig zijn, pas kunnen ontspannen als iets af is, het zelf willen doen, maar ook het maken van eigen keuzes. Wat echter nog belangrijker is, is jullie onvoorwaardelijke liefde, het vertrouwen dat jullie geven, en het plezier dat jullie met ons blijven maken. Lieve familie, jullie zijn de warme basis waar ik altijd op kan terugvallen! *How wonderful life is while you're in the world!*

Lieve Melissa, mijn dakduif, snorlex, pipoman, snoggelman, babymachine, of hoe ik je wel eens noem, ik ben enorm gelukkig met jou. Ik begrijp soms niet wat jij leuk vindt aan iemand die als een onrustige, hardlopende, buiten-douchende, vroeg-opstaande GFT-bak door het leven gaat. Toch bied je altijd een luisterend oor en heb je altijd het geduld om mij mijn gang te laten gaan, en dat waardeer ik ontzettend. Ik heb enorm genoten van onze reisjes naar Slovenië, IJsland, Canada, en Argentinië van afgelopen jaren, en ik hoop er, samen met ons komende kleintje, nog vele te mogen maken. Ik kan niet wachten. Dankjewel dat je er altijd voor mij bent. Ik hou van jou/jullie!

Curriculum Vitae/About the author

Mitchel Arno Molenaar was born on December 11, 1994, in Schagen, The Netherlands. He completed his secondary education at Regius College, Schagen, in 2013. After Mitchel obtained his Propedeuse in Medical Natural Sciences at the Vrije Universiteit Amsterdam, Mitchel enrolled the new study Technical Medicine, jointly offered by Delft University of Technology, Erasmus Medical Center, and Leiden University Medical Center. During the Technical Medicine master's program, Mitchel gained clinical and technical experience through multiple internships at renowned institutions in the Netherlands and abroad. He taught himself Python and enrolled in extracurricular courses in machine learning. For his master's thesis, Mitchel conducted research in the Department of Sports Medicine at the Erasmus Medical Center, where he deployed deep learning to assess lower limb function in patients with patellar tendinopathy. He was one of the first ten graduates of the program.



After graduation, Mitchel aspired to further specialize in machine learning with the goal of contributing to healthcare. He actively sought positions where he could make a meaningful impact, particularly in the field of cardiovascular diseases, given their profound effect on public health.

In 2020, he started his PhD journey at the Department of Cardiology at Amsterdam University Medical Center (AUMC), under the supervision of Prof. dr. S.A.J. Chamuleau. During his PhD, Mitchel collaborated with other departments on AI within the innovation team. He developed a data pipeline for the collection of clinical and imaging data, formed strong partnerships with Philips, and mentored a total of 34 students over the course of his PhD. He was among the pioneers in building machine learning models using AUMC data. His work led to invitations to present at both national and international conferences. After completing his PhD, Mitchel started working as Data Scientist at Thuisarts, where he applies his knowledge to enhance public health literacy.

Outside of his academic pursuits, Mitchel resides with his girlfriend, Melissa, in Waddinxveen. In his free time, Mitchel enjoys the great outdoors, and is an enthusiastic runner, achieving marathon times under three hours. In addition, he really enjoys watching Kees van der Spek expose and confront the Yahoo boys. Based on his "impressive" CV, you wouldn't expect it, but he remains a remarkably modest person.