Automated echocardiographic detection of severe coronary artery disease using artificial intelligence

Upton, Ross, Mumith, Angela, Beqiri, Arian, Parker, Andrew, Hawkes, William, Gao, Shan, Porumb, Mihaela, Sarwar, Rizwan, Marques, Patricia, Markham, Deborah, Kenworthy, Jake, O'Driscoll, Jamie M, Hassanali, Neelam, Groves, Kate, Dockerill, Cameron, Woodward, William, Alsharqi, Maryam, McCourt, Annabelle, Wilkes, Edmund H, Heitner, Stephen B, Yadava, Mrinal, Stojanovski, David, Lamata, Pablo, Woodward, Gary and Leeson, Paul

This is the accepted version of:

Ross Upton, Angela Mumith, Arian Beqiri, Andrew Parker, William Hawkes, Shan Gao, Mihaela Porumb, Rizwan Sarwar, Patricia Marques, Deborah Markham, Jake Kenworthy, Jamie M. O’Driscoll, Neelam Hassanali, Kate Groves, Cameron Dockerill, William Woodward, Maryam Alsharqi, Annabelle McCourt, Edmund H. Wilkes, Stephen B. Heitner, Mrinal Yadava, David Stojanovski, Pablo Lamata, Gary Woodward, Paul Leeson,

Automated Echocardiographic Detection of Severe Coronary Artery Disease Using Artificial Intelligence,

JACC: Cardiovascular Imaging,

2021,

,

ISSN 1936-878X,

https://doi.org/10.1016/j.jcmg.2021.10.013.

(https://www.sciencedirect.com/science/article/pii/S1936878X21007804)

Automated Echocardiographic Detection of Severe Coronary Artery Disease using Artificial Intelligence

Brief title: Stress echocardiography and artificial intelligence

Ross Upton, PhD,[1,2] Angela Mumith, PhD*,[1] Arian Beqiri*, PhD,[1] Andrew Parker*, PhD,[1] William Hawkes*, PhD,[1] Shan Gao, PhD,[1] Mihaela Porumb, PhD,[1] Rizwan Sarwar, MD[2,] Patricia Marques, BSc,[1] Deborah Markham, PhD,[1], Jake Kenworthy, Bsc,[1] Jamie M. O'Driscoll, PhD,[1,3] Neelam Hassanali, PhD,[1] Kate Groves, PhD,[1] Cameron Dockerill, BSc,[2] William Woodward, BSc,[2] Maryam Alsharqi, MSc,[2] Annabelle McCourt, MSc,[2] Edmund H. Wilkes, PhD,[1] Stephen B. Heitner, MD,[4] Mrinal Yadava, MD,[4] David Stojanovski, MEng,[5] Pablo Lamata, PhD,[5] Gary Woodward, PhD[1], Paul Leeson, MB BChir PhD, [1,2]

* These authors contributed equally to the manuscript

[1] Ultromics Ltd, Oxford, UK. [2] Cardiovascular Clinical Research Facility, RDM Division of Cardiovascular Medicine, University of Oxford, Oxford, UK. [3] School of Human and Life Sciences, Canterbury Christ Church University, Kent, UK [4] Knight Cardiovascular Institute, Oregon Health & Science University, Portland Oregon, USA. [5] Dept. of Imaging Sciences and Biomedical Engineering, King's College London, London, UK.

Funding: Data and images used in the training dataset were collected with support of NIHR HEE Healthcare Science Research Fellowship [NIHR-HCS-P13-04-001]; Cardiovascular Clinical Research Facility, University of Oxford; Ultromics Ltd.; Lantheus Medical Imaging Inc., NIHR Oxford Biomedical Research Centre, University of Oxford and Oxford BHF Centre for Research Excellence. PLa holds a Wellcome Trust Senior Research Fellowship (209450/Z/17/Z).

Disclosures: RU is the CEO, co-founder, and a shareholder of Ultromics Ltd, which develops AI echocardiography software. PLe is a co-founder, a shareholder and non-executive director of Ultromics Ltd; has previously consulted for Intelligent Ultrasound and has held research grants from Lantheus Medical Imaging. RU, PLe, DM and EW are inventors on filed patents in the field of echocardiography. AM, AB, AP, WH, SG, MP, PM, DM, JK, JOD, NH, KG and GW are employees of Ultromics Ltd. The remaining authors have nothing to disclose.

Corresponding author: Professor Paul Leeson, Oxford Cardiovascular Clinical Research Facility, RDM Division of Cardiovascular Medicine, John Radcliffe Hospital, Oxford. OX3 9DU. Tel: +44 1865 226829 paul.leeson@cardiov.ox.ac.uk

Abstract

Background: Coronary artery disease is the leading global cause of mortality and morbidity and stress echocardiography remains one of the most commonly used diagnostic imaging tests.
Objectives: To establish whether an artificially intelligent system can be developed to automate stress echocardiography analysis and support clinician interpretation.
Methods: An automated image processing pipeline was developed to extract novel geometric and kinematic features from stress echocardiograms collected as part of a large, UK-based prospective, multi-centre, multi-vendor study. An ensemble machine learning classifier was trained, using the extracted features, to identify patients with severe coronary artery disease on invasive coronary angiography. The model was tested in an independent US study. How availability of an AI classification might impact clinical interpretation of stress echocardiograms was evaluated in a randomised cross-over reader study.
Results: Acceptable classification accuracy for identification of patients with severe coronary artery disease in the training dataset was achieved on cross fold validation based on 31 unique geometric and kinematic features, with a specificity of 92.7% and a sensitivity of 84.4%. This accuracy was maintained in the independent validation dataset. The use of the AI classification tool by clinicians increased inter-reader agreement and confidence as well as sensitivity for detection of disease by 10% to achieve an AUROC of 0.93.
Conclusion: Automated analysis of stress echocardiograms is possible using artificial intelligence and provision of automated classifications to clinicians when reading stress echocardiograms could improve accuracy, inter-reader agreement and reader confidence.

Key words: Stress echocardiography; artificial intelligence; coronary artery disease;

Abbreviations:
A2C: Apical 2 Chamber view
A3C: Apical 3 Chamber view
A4C: Apical 4 Chamber view
CAD: Coronary Artery Disease
CNN: Convolutional Neural Network
EVAREST: Echocardiography Value and Accuracy at REst and Stress
ICA: Invasive Coronary Angiography
MRMC: Multiple Reader Multiple Case
SAX: Parasternal Short-Axis Mid-Ventricular view
SE: Stress Echocardiography

Introduction

Coronary artery disease is a major cause of morbidity and mortality worldwide[1,2] and effective management relies on cardiac imaging to risk-stratify those who may require further treatment[3,4]. Stress echocardiography (SE) is one of the most widely used modalities for non-invasive assessment of coronary artery disease due to its low cost, absence of ionising radiation and high patient tolerability[3–5]. The test requires a clinician to compare images of left ventricular wall motion acquired before and after exercise or pharmacological stress. Myocardial contractility increases as heart rate rises with stress but in areas affected by flow-limiting coronary artery disease, the myocardium becomes ischaemic and contractility reduces[6]. This resultant regional wall motion abnormality needs to be identified 'by eye' and without adequate training and experience, inter- and intra-observer variability can be high.

Automation of image assessment could overcome this limitation, increase confidence and broaden use. Recent application of artificial intelligence (AI) to image analysis, including of echocardiograms has demonstrated automation of tasks previously reliant on expert opinion is possible[7,8]. However, translation into clinical practice has been slow because of requirements for robust, generalizable systems and concerns about how AI-derived information might impact clinician performance. We developed a bespoke sequence of AI algorithms that could automatically process and extract novel imaging features from stress echocardiograms. We then incorporated the novel features into a machine learning model trained to identify patients with significant coronary artery disease. The performance of this model was tested in an independent dataset and we then performed a randomised, blinded trial to assess how the provision of an AI classifier impacts clinician interpretation of stress echocardiograms.

Methods

Model development and validation

Training and validation dataset

Model development and validation was based on a dataset of clinical information and images extracted from the multicentre EVAREST study (Echocardiography Value and Accuracy at REst and Stress, ClinicalTrials.gov Identifier: NCT03674255). Images had been acquired from hospitals with a range of sizes, type of operators and ultrasound vendor equipment representative of “real world” stress echocardiography. Recruitment started in July 2015 and is ongoing with data and images stored within the core laboratory at the Cardiovascular Clinical Research Facility (CCRF), University of Oxford. Participants gave informed consent for clinical follow up via medical records for up to 10 years after recruitment. Ethical approval for the EVAREST study was obtained from Health Research Authority NRES Committee South Central – Berkshire (IRAS reference: 14/SC/1437). For model training, a set of images were identified that: included apical-4-chamber (A4C), apical-2-chamber (A2C) and parasternal short-axis mid-ventricular (SAX) views at rest and stress; had endocardial visualization in at least 14 of 16 segments in available images (based on consensus review by three 3 BSE accredited cardiac physiologists); had end-diastolic (ED) and end-systolic (ES) frames with a minimum of 4 frames between ED and ES; were a diagnostic stress echocardiogram (target heart rate or double product reached during study or appropriate end point reached for diagnostic interpretation); no previous history of coronary artery bypass graft (CABG) or other cardiac surgery.

Feature extraction

To extract image features for training the model, all the image datasets were segmented and contoured using a bespoke, fully automated AI pipeline (full details are provided in Supplemental Methods and Data). In brief, views are classified based on a 2D convolutional neural network (CNN)[9–12] that also identifies whether they are contrast or non-contrast images. Depending on the identified view, images are then passed through one of three auto-contouring CNNs (one for A2C/A4C contrast data, one for A2C/A4C non-contrast and one for SAX data) to first segment, and then contour, the LV endocardial border on each image frame. From these contours at, and between, ED and ES, multiple features were generated, including both (1) routine clinical measures such as ejection fraction and global longitudinal strain[13] as well as (2) novel features engineered specifically for this stress echocardiography project based on prior expert knowledge of myocardial wall motion characteristics and previously reported methods for defining geometric (shape) and kinematic (mechanical, rate) changes. In total, around 7000 features were developed to describe global and regional contour shape and temporal deformation. Examples of features that were included in the final model are reported in Supplemental Table 3.

Model training and validation

A supervised machine learning model based on the novel image-derived features was trained using Python 3.6 with the scikit-learn library to identify patients with severe coronary disease. Presence of severe coronary disease was based on data from invasive coronary angiography (ICA) whether performed electively or following acute admission to hospital within six months of the SE. In the case of repeated ICA the first instance was used as the reference. The presence of severe disease was assigned if ≥50% stenosis was evident on ICA in the left main stem (LMS) or ≥70% in the proximal to mid left anterior descending artery (LAD), proximal left circumflex (LCx) or proximal to mid right coronary artery (RCA). All other patients i.e. those with <70% stenosis on ICA or in whom ICA was not considered clinically necessary were classed as not having ‘severe coronary disease’. Disease classification was determined by an adjudication committee, comprising of at least one board-certified cardiologist, blinded to the SE result.

Three classifiers were trained (support-vector machine, random forest and logistic regression) and collated to create an ensemble classifier using a soft voting strategy. Stratified and repeated 2-fold nested cross validation (CV) was adopted[14] with hyperparameter optimisation. Imaging feature selection for each fold and repeat was performed in the training fold of the outer loop. Imaging feature selection and hyperparameter optimisation was carried out 20 times over 2 folds and 10 repeats. Feature selection relied on the Boruta multivariate approach[15] and the top selected features were then included in the final classifier model. Model performance was assessed as an overall metric following nested CV.

To better understand the clinical relevance of the selected features, post-hoc feature analysis was performed. Each of the selected features was evaluated using ROC curve analysis against adjudicated disease classification and each feature was evaluated using the optimum threshold based on the maximum Youden's Index. In addition, regionality of derived features was examined to understand whether they aligned with clinical interpretation of the stress echocardiogram regional wall motion abnormality.

Independent testing

Once the model had been developed it was independently tested in a retrospective study of patients who had undergone stress echocardiography at Oregon Health Science University Hospital (OHSU; Portland Oregon) (RAINIER study). Applicable stress echocardiogram images and related data was collected from studies performed between 2011 and 2017. All images were required to have been deemed appropriate for clinical diagnosis by the clinician performing the study. All patients had follow up to at least six months through clinical record review and adjudication of disease severity using the same process as for the training dataset, undertaken by an adjudication committee blinded to results of the SE. Datasets were collected in a consecutive 'case' (evidence of severe coronary disease) – 'control' manner until the predetermined sample size was achieved. An Institutional Review Board (IRB) waiver for consent was obtained based on the anonymised retrospective nature of the study. All images were processed using the same automated AI pipeline and then the model was used to identify patients which it classed as having ‘severe coronary disease’. Accuracy of the model was compared against the adjudicated disease classification.

Reader study

The impact of provision of an AI-derived disease classifier on clinician interpretation of the stress echocardiogram was studied using a multiple reader multiple case (MRMC) randomised cross-over design[16,17]. Two US-based (accredited by the American Society of Echocardiology) and two UK-based (accredited by the British Society of Echocardiography) physicians/echocardiographers who were independent of any other part of the investigation and had at least 2 years experience of stress echocardiogram interpretation participated[18,19]. Reader experience ranged from ~350 stress echo per year to 2 years transthoracic echocardiography with trainee experience in stress echocardiography. Readers were presented with all the stress echo studies used for the RAINIER validation trial and asked to identify patients who, in their expert opinion, had severe coronary disease, as defined for training the machine learning model. The study design meant readers saw the images twice. The first time 50% were accompanied by a report containing information on whether the AI had classified the patient as having severe coronary disease (Supplemental Figure 7). Readers were aware the AI-based classification was not 100% accurate and were free to choose whether to use the classification in their clinical interpretation. Studies were presented in a randomised order and after a one-month washout period all readers were shown the images again but this time with the AI report provided for the other 50% of studies. After each study, the readers were also asked to provide a binary (confident or not confident) measure of their confidence in their clinical interpretation of the study.

Statistical Analysis

Previously observed cross-validation performance metrics were used to estimate the sample size required for this study. The reader study was powered and performed using the FDA iMRMC application. Data from a pilot study informed the sample size with 100 normal and 50 CAD patients providing 80% power to detect a 0.042 difference in AUROC, using four readers. Diagnostic performance was assessed by sensitivity, specificity, ROC and performance recall analyses and t-distribution tests (T-Tests) and standard error measurements were calculated using Brandon D Gallas variance components[20]. Standard approaches for ROC curve generation in a MRMC analysis[16,17] were used based on sensitivity and specificity calculated at four thresholds determined by the diagnostic confidence level reported by the reader: 1- confident positive read 2 – not-confident/probably positive read, 3 – not confident/probably negative read and 4 – confident negative read. Separate ROC curves were generated for reads with and without AI assistance to determine differences in area under the curves. In addition, McNemar's test was used to assess change in confidence with use of AI interpretation. Continuous variables were expressed as mean (±SD) or median (interquartile range (IQR)) according to data distribution and compared using the Student t test or Wilcoxon ranks sum test, as appropriate. Categorical data, presented as number and percentages, were compared using χ^2 test.

Results

Study Populations

The demographics of the datasets are described in Table 1. In summary, 578 patient datasets from the EVAREST study were used for AI training, of which 58% were male with a mean BMI of 30 kg/m^2, 40% had hypertension and 31% had a previous PCI. Severe coronary disease was present on ICA in 14.7%. Two thirds of images were from Philips machines and one third from GE. In total, 87.2% were dobutamine stress studies and 12.8% were exercise stress. The independent RAINIER testing dataset comprised of 154 studies, 50% were male with a mean BMI of 30 kg/m^2, 62.1% had hypertension and 8.4% had a previous PCI. Twenty one had resting wall motion abnormalities, of which, 11 patients had abnormalities in the anterior/anterolateral wall, 11 in the inferior/inferolateral wall, 6 in the septum and 16 had involvement of the apex. Severe coronary disease on ICA was present in 29.2%. 96.1% of images were from Philips machines and 3.9% from GE. In total, 35.7% were dobutamine stress studies and 64.3% were exercise stress.

Development and Performance of Machine Learning Classifier

The feature selection process identified 31 features out of 6,748 novel image-derived features, which contributed to disease classification (Supplemental Table 3). Of these features, 20 were derived from the A4C view, 2 from the A2C view and 9 from the SAX view, with apical lateral and mid anterolateral sections being chosen most frequently during feature selection. Out of the 31 selected features, 15 were markers of the magnitude of regional wall motion abnormality and 16 were markers of endocardial velocity, or tardokinesis[5]. Visual inspection of the myocardial regions from which features were used for classification were broadly similar to those regions identified by clinicians as having regional wall motion abnormalities during the clinical reading of the stress echocardiogram (Figure 1 and Supplemental Table 2), although the model was less likely to incorporate features from the basal segments. Evaluation of individual features demonstrated that they were each moderately effective for identification of patients with severe disease (AUROC range 0.760 – 0.867, see Supplemental Table 3 and Figure 2). There was also evidence of stratification of disease clusters when two features were used in bi-variate plots. When features were combined with hyperparameter optimisation in the ensemble machine learning classifier during a 10 repeat 2-fold CV, an AUROC of 0.934 (Figure 2) was present with an optimal specificity of 85.7% (95% CI 82.7%, 88.9%) and sensitivity of 86.7% (95% CI 80.2%, 94.3%). Independent testing of this ensemble classifier in the RAINIER study testing dataset produced a similar AUROC of 0.927 (Figure 2) with a specificity of 92.7% (95% CI 87.8%, 97.6%) and a sensitivity of 84.4% (95% CI 73.9%, 95.0%) at the classification threshold set in the training dataset. In a subgroup analysis excluding those with known coronary artery disease or resting wall motion abnormalities sensitivity remained at 90.5% and specificity 88.4%.

Clinical Performance with and without AI Assistance

When provided with the AI-based classification to assist in their interpretation of the RAINIER study image datasets, all readers exhibited an increase in mean ± SD sensitivity (Figure 3) from 85.0 ± 4.0 % to 95.0 ± 3.0% (Δsensitivity = + 10.0, 95% CI 6.5, 13.5, $p<0.05$). Mean reader specificity remained consistent from 83.6 ± 7.9 % without to 85.0 ± 3.2 % with the assistance of the AI-based classifier (Δspecificity = + 1.4 %, 95% CI -5.7, 8.5, $p>0.05$, Supplemental Table 4). Reads with the AI classifier resulted in a 10% increase in number of confident reads (440 unassisted vs 483 assisted) and a corresponding 29% decrease in non-confident reads (152 unassisted vs 109 assisted) ($p<0.001$). To evaluate how reader accuracy and confidence changed when assisted by AI-based classification, stress echo interpretation and confidence ratings were compounded to construct MRMC ROC curves. Readers exhibited significant increases in the mean ± standard error AUROC from 0.877 ± 0.019 without to 0.931 ± 0.028 with the assistance of the classifier (ΔAUROC = +0.054, 95% CI 0.032, 0.077, $p<0.05$, Figure 3). Indeed, with the assistance of the AI-based classifier, the AUROC of 2 readers exceeded the performance of the AI classifier on its own (AUROC = 0.927), with Reader 3 achieving 100% sensitivity (Supplemental table 4). To further evaluate the impact of the AI-based classification on reader performance the level of agreement between the four readers was then evaluated (Figure 3). This demonstrated in all reader comparisons, reader agreement improved from between 0.68 to 0.79 up to 0.83 to 0.97. To determine whether presence of resting wall motion abnormalities had an impact on performance the analysis was repeated just on those with normal wall motion at rest. There was no difference in performance of either the AI-only model or the AI-assisted reads [see Supplemental Tables 5 and 6].

Discussion

This study demonstrates that an artificially intelligent system can automatically 'read' stress echocardiograms, without the input of a clinician, and differentiate patients who may require revascularisation if they have angiography from those who may be better managed medically, as they are not likely to have severe coronary disease on angiography or are unlikely to have an acute cardiac event during follow up. The AI alone achieved very good diagnostic accuracy but to increase acceptability of AI during adoption into clinical practice it is likely clinicians will initially incorporate information from AI into their decision making. Therefore, we performed a reader study, and demonstrated that provision of an AI classifier result to a clinician improved both their performance and their confidence in diagnosis. These results demonstrate a potential for automated AI-based methodologies to augment clinical performance in stress echocardiography[21–24].

A series of AI innovations were required to achieve this automation of stress echo interpretation. Central to the image analysis pipeline are LV-contouring CNNs capable of tracking the endocardial border effectively in greater than 90% of cases. When our project commenced, no solutions for this had been published but, in recent years, automated segmentation and contouring of echocardiography has become accepted. The approach we developed is similar to other reported segmentation approaches for routine echocardiography[25–28]. Additionally, we trained networks to handle contrast-enhanced images across the broad range of heart rates typical in stress echocardiography. To achieve this we used multi-vendor datasets and completed manual ground truth contouring. Use of images from stress echocardiogram protocols in routine use within multiple sites in the UK and US were selected to ensure algorithm training and testing would be applicable to “real world” variations in image acquisition, operators and clinical protocols. Inclusion of image datasets acquired with both pharmacological and exercise stress also maximised clinical generalisability[29,30]. In this study, image quality needed to be considered of diagnostic quality by the operator. Whether image quality limits differ with AI interpretation could be explored further with experimental inclusion of a broad range of images of varying quality.

To achieve optimal performance we considered three supervised machine learning classifiers[31,32]. This approach was selected over a deep learning classification model for two reasons[33]. Firstly, we could use clinical expertise to engineer a range of features expected to capture novel echocardiographic and myocardial patterns predictive of coronary disease. Secondly, post-hoc analysis was used to generate insights into how individual global and regional features contribute to the disease classification model. This provided an internal ‘sense check’ that regions within the echocardiogram being used by the AI for diagnosis made clinical sense, while identifying several novel geometric and kinematic features highly predictive of disease. While most features were derived from peak stress images, several resting features were included, raising the possibility that a proportion of patients may be classifiable based on resting echocardiograms alone.

A key element of the study was the independent testing and clinical trialling of the trained model. Clinical adoption of AI within cardiology requires the clinician to take information derived from automated analysis and incorporate it into their clinical decision making. It is possible an AI tool, if inappropriately ignored, or accepted, by a clinician, could have a detrimental impact on patient outcome[34]. We therefore undertook a randomised crossover design, including a month washout period between reads to ensure blinding of the reader, to understand how use of an AI diagnostic aid influences decision making. Readers had different levels of prior experience of stress echocardiography and the randomised crossover design minimised potential training effects of the AI on the reader. We were able to demonstrate in an unbiased fashion that both performance and confidence were higher in scans reviewed with the assistance of the AI diagnostic tool resulting in a 10% increase in sensitivity[35]. We did not, however, see variation with level of operator expertise but this could be explored further in a larger study. Further, prospective randomised controlled trials will be of value to understand the impact of AI on patient outcome.

Study limitations

This study has several limitations. Firstly, we have not compared against another diagnostic test i.e. stress echo against a second diagnostic test e.g. FFR. Therefore, the model cannot exclude the possibility of some degree of disease in those patients who did not undergo angiography. However, based on the follow up data, we know these patients did not have acute cardiac events and therefore were appropriate to manage medically. This classification is consistent with recent randomised trials such as (PROMISE[36], SCOT-HEART[37], CE-MARC2[38], FORECAST[39], ISCHAEMIA[40]) that focus on how imaging influences clinical practice and outcome, including highlighting that routine referral for angiography may not be warranted for many patients whom it would be better to manage medically after their imaging test. The next phase of our work is a randomised controlled trial (PROspective randomised controlled Trial Evaluating the Use of artificial intelligence in Stress echocardiography (PROTEUeS, ISRCTN registry ID ISRCTN15113915), which will formally test whether provision of this AI-derived guidance impacts clinical outcome and resource utilisation, such as angiography, for the patient. Secondly, the training sample size was relatively small and to avoid biased overestimations of summary performance statistics, missing and inconclusive data were handled using routine approaches[41,42]. In this manner, all cases were included as far as practicable to minimize associated biases. To ensure this did not lead to overestimation of performance the stability of the model was tested in the independent dataset. The dataset used for testing also varied in frequency of clinical characteristics from the training dataset consistent with approaches to ensure robust, generalizable independent testing datasets[43] Thirdly, our disease classification of 'severe coronary disease' was based on clinician interpretation of invasive coronary angiography and we did not have access to quantitative measures of coronary stenosis to confirm severity. Although we used an adjudication committee blinded to the SE result to confirm diagnosis the imprecision of stenosis assessment might have reduced accuracy of training. Fourthly, the model was trained to identify severe coronary disease as a 'yes/no' classification. Information on angiography was available in all patients in the testing dataset who had adverse events after stress echocardiography. However, some degree of coronary disease is not excluded in those classified as 'non-severe coronary disease'. In clinical application this group would require clinical assessment to decide on need for most appropriate management. In the future, it may be possible to develop and train further models to provide classification of disease severities. Fifthy, we did not differentiate between mode of stress and future work may be of value to understand whether bespoke models for each stressor could increase accuracy. Finally, we also did not take account of ethnicity or race in development of the model and further work could be considered to understand whether incorporating this information into models could optimise them further.

Conclusions

In conclusion, we have demonstrated that an artificially intelligent system can be developed to autonomously 'read' typical stress echocardiograms currently being acquired for clinical diagnosis and differentiate patients likely to have severe coronary disease on angiography from those who will not have severe angiographic coronary disease and/or are at low risk of an imminent cardiac event. We have also shown that when a clinician is provided with this AI interpretation and asked to make a clinical decision, they are more accurate and more confident in their decision. Such AI technologies could have the potential to significantly impact clinical workflows and patient care, particularly with regard selection of patients for invasive testing. Further work is now needed to prospectively evaluate these tools in formal randomised trials to determine their impact on patient outcomes.

Clinical perspectives

Clinical competencies

Stress echocardiography is one of the most commonly used imaging tests to diagnose coronary artery disease and decide whether patients need further investigation or treatment with invasive coronary angiography. The interpreting clinician needs to identify regional wall motion abnormalities on the echocardiogram with the eye and therefore can vary depending on their expertise.

Translational outlook

(1) This study shows it is possible to use artificial intelligence to automatically process stress echocardiograms and classify whether the patient is likely to have severe coronary disease, which might need further investigation with an angiogram

(2) Furthermore, if clinicians are provided with the automatic classification when they are looking at the images it reduces variability amongst readers and increases both their accuracy and confidence in diagnosis.

References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Ahmed M, Aksut B, Alam T, Alam K, Alla F, Alvis-Guzman N, Amrock S, Ansari H, Arnlov J, Asayesh H, Atey TM, Avila-Burgos L, Awasthi A, Banerjee A, Barac A, Barnighausen T, Barregard L, Bedi N, Belay Ketema E, Bennett D, Berhe G, Bhutta Z, Bitew S, Carapetis J, Carrero JJ, Malta DC, Castaneda-Orjuela CA, Castillo-Rivas J, Catala-Lopez F, Choi J-Y, Christensen H, Cirillo M, Cooper LJ, Criqui M, Cundiff D, Damasceno A, Dandona L, Dandona R, Davletov K, Dharmaratne S, Dorairaj P, Dubey M, Ehrenkranz R, El Sayed Zaki M, Faraon EJA, Esteghamati A, Farid T, Farvid M, Feigin V, Ding EL, Fowkes G, Gebrehiwot T, Gillum R, Gold A, Gona P, Gupta R, Habtewold TD, Hafezi-Nejad N, Hailu T, Hailu GB, Hankey G, Hassen HY, Abate KH, Havmoeller R, Hay SI, Horino M, Hotez PJ, Jacobsen K, James S, Javanbakht M, Jeemon P, John D, Jonas J, Kalkonde Y, Karimkhani C, Kasaeian A, Khader Y, Khan A, Khang Y-H, Khera S, Khoja AT, Khubchandani J, Kim D, Kolte D, Kosen S, Krohn KJ, Kumar GA, Kwan GF, Lal DK, Larsson A, Linn S, Lopez A, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70:1–25.
2. Nowbar A, Mauro G, Howard J, Francis D, Rasha A-L. Mortality From Ischemic Heart Disease. Circ Cardiovasc Qual Outcomes [Internet]. 2019;12:e005375. Available from: https://doi.org/10.1161/CIRCOUTCOMES.118.005375
3. Wolk MJ, Bailey SR, Doherty JU, Douglas PS, Hendel RC, Kramer CM, Min JK, Patel MR, Rosenbaum L, Shaw LJ, Stainback RF, Allen JM. ACCF/AHA/ASE/ASNC/HFSA/HRS/SCAI/SCCT/SCMR/STS 2013 multimodality appropriate use criteria for the detection and risk assessment of stable ischemic heart disease: a report of the American College of Cardiology Foundation Appropriate Use

Criteria Task Forc. J Am Coll Cardiol. 2014;63:380–406.

4. Saraste A, Barbato E, Capodanno D, Edvardsen T, Prescott E, Achenbach S, Bax JJ, Wijns W, Knuuti J. Imaging in ESC clinical guidelines: chronic coronary syndromes. Eur Heart J Cardiovasc Imaging. 2019;20:1187–1197.

5. Pellikka PA, Arruda-Olson A, Chaudhry FA, Chen MH, Marshall JE, Porter TR, Sawada SG. Guidelines for Performance, Interpretation, and Application of Stress Echocardiography in Ischemic Heart Disease: From the American Society of Echocardiography. J Am Soc Echocardiogr. 2020;33:1-41.e8.

6. Nesto RW, Kowalchuk GJ. The ischemic cascade: temporal sequence of hemodynamic, electrocardiographic and symptomatic expressions of ischemia. Am J Cardiol. 1987;59:23C-30C.

7. Massalha S, Clarkin O, Thornhill R, Wells G, Chow BJW. Decision Support Tools, Systems, and Artificial Intelligence in Cardiac Imaging. Can J Cardiol [Internet]. 2018 [cited 2019 Sep 11];34:827–838. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0828282X18303696

8. Alsharqi M, Woodward J, Mumith A, Markham C, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract [Internet]. 2018;5:R115–R125. Available from: https://erp.bioscientifica.com/view/journals/echo/5/4/ERP-18-0056.xml

9. Gao X, Li W, Loomes M, Wang L. A fused deep learning architecture for viewpoint classification of echocardiography. Inf Fusion. 2017;36:103–113.

10. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan CR, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fast and accurate view classification of echocardiograms using deep learning (Supplemental Material). :2–6.

11. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. npj Digit Med. 2018;1.
12. Kusunose K, Haga A, Inoue M, Fukuda D, Yamada H, Sata M. Clinically feasible and accurate view classification of echocardiographic images using deep learning. Biomolecules. 2020;10:1–8.
13. Aloysius N, Geetha M. A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing (ICCSP). 2017. p. 588–592.
14. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7:1–8.
15. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw [Internet]. 2010;36:1–13. Available from: http://www.jstatsoft.org/v36/i11/paper
16. Dendumrongsup T, Plumb AA, Halligan S, Fanshawe TR, Altman DG, Mallett S. Multi-reader multi-case studies using the area under the receiver operator characteristic curve as a measure of diagnostic accuracy: systematic review with a focus on quality of data reporting. PLoS One. 2014;9:e116018.
17. Hadjiiski L, Chan H-P, Sahiner B, Helvie MA, Roubidoux MA. Quasi-continuous and discrete confidence rating scales for observer performance studies: Effects on ROC analysis. Acad Radiol [Internet]. 2007;14:38–48. Available from: https://pubmed.ncbi.nlm.nih.gov/17178364
18. Pellikka PA, Nagueh SF, Elhendy AA, Kuehl CA, Sawada SG. American Society of Echocardiography Recommendations for Performance, Interpretation, and Application of Stress Echocardiography. J Am Soc Echocardiogr [Internet]. 2007;20:1021–1041. Available from: http://www.sciencedirect.com/science/article/pii/S0894731707005378
19. Picano E, Lattanzi F, Orlandini A, Marini C, L'Abbate A. Stress echocardiography and

the human factor: The importance of being expert. J Am Coll Cardiol. 1991;17:666–669.

20. Gallas BD, Bandos A, Samuelson FW, Wagner RF. A Framework for Random-Effects ROC Analysis: Biases with the Bootstrap and Other Variance Estimators. Commun Stat - Theory Methods [Internet]. 2009;38:2586–2603. Available from: https://doi.org/10.1080/03610920802610084
21. Uusitalo V, Luotolahti M, Pietila M, Wendelin-Saarenhovi M, Hartiala J, Saraste M, Knuuti J, Saraste A. Two-Dimensional Speckle-Tracking during Dobutamine Stress Echocardiography in the Detection of Myocardial Ischemia in Patients with Suspected Coronary Artery Disease. J Am Soc Echocardiogr. 2016;29:470-479.e3.
22. Rumbinaite E, Zaliaduonyte-Peksiene D, Lapinskas T, Zvirblyte R, Karuzas A, Jonauskiene I, Viezelis M, Ceponiene I, Gustiene O, Slapikas R, Vaskelyte JJ. Early and late diastolic strain rate vs global longitudinal strain at rest and during dobutamine stress for the assessment of significant coronary artery stenosis in patients with a moderate and high probability of coronary artery disease. Echocardiography [Internet]. 2016;33:1512–1522. Available from: https://doi.org/10.1111/echo.13282
23. Yang L-T, Kado Y, Nagata Y, Otani K, Otsuji Y, Takeuchi M. Strain Imaging with a Bull’s-Eye Map for Detecting Significant Coronary Stenosis during Dobutamine Stress Echocardiography. J Am Soc Echocardiogr. 2017;30:159-167.e1.
24. Mansour MJ, AlJaroudi W, Hamoui O, Chaaban S, Chammas E. Multimodality imaging for evaluation of chest pain using strain analysis at rest and peak exercise. Echocardiography [Internet]. 2018;35:1157–1163. Available from: https://doi.org/10.1111/echo.13885
25. Dias JMB, Leitào JMN. Wall position and thickness estimation from sequences of echocardiographic images. IEEE Trans Med Imaging. 1996;15:25–38.

26. Jacob G, Noble JA, Mulet-Parada M, Blake A. Evaluating a robust contour tracker on echocardiographic sequences. Med Image Anal. 1999;3:63–75.

27. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging. 2006;25:987–1010.

28. Azarmehr N, Ye X, Sacchi S, Howard JP, Francis DP, Zolgharni M. Segmentation of Left Ventricle in 2D Echocardiography Using Deep Learning. In: Communications in Computer and Information Science. 2020. p. 497–504.

29. Sacristan JA, Dilla T. No big data without small data: learning health care systems begin and end with the individual patient. J Eval Clin Pract. 2015;21:1014–1017.

30. Hu G, Peng X, Yang Y, Hospedales TM, Verbeek J. Frankenstein: Learning Deep Face Representations Using Small Data. IEEE Trans Image Process. 2018;27:293–303.

31. Géron A. Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems. 2017.

32. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online [Internet]. 2014;13:94. Available from: http://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-13-94

33. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan CR, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018;138:1623–1635.

34. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, Marwick TH. Artificial Intelligence in Cardiovascular Imaging. J Am Coll Cardiol [Internet]. 2019;73:1317 LP – 1335. Available from: http://www.onlinejacc.org/content/73/11/1317.abstract

35. Fihn SD, Blankenship JC, Alexander KP, Bittl JA, Byrne JG, Fletcher BJ, Fonarow GC, Lange RA, Levine GN, Maddox TM, Naidu SS, Ohman EM, Smith PK. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, . J Am Coll Cardiol. 2014;64:1929–1949.

36. Douglas PS, Hoffmann U, Patel MR, Mark DB, Al-Khalidi HR, Cavanaugh B, Cole J, Dolor RJ, Fordyce CB, Huang M, Khan MA, Kosinski AS, Krucoff MW, Malhotra V, Picard MH, Udelson JE, Velazquez EJ, Yow E, Cooper LS, Lee KL. Outcomes of Anatomical versus Functional Testing for Coronary Artery Disease. N Engl J Med [Internet]. 2015;372:1291–1300. Available from: https://doi.org/10.1056/NEJMoa1415516

37. SCOT-HEART. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. N Engl J Med [Internet]. 2018;379:924–933. Available from: https://doi.org/10.1056/NEJMoa1805971

38. Greenwood JP, Ripley DP, Berry C, McCann GP, Plein S, Bucciarelli-Ducci C, Dall'Armellina E, Prasad A, Bijsterveld P, Foley JR, Mangion K, Sculpher M, Walker S, Everett CC, Cairns DA, Sharples LD, Brown JM. Effect of Care Guided by Cardiovascular Magnetic Resonance, Myocardial Perfusion Scintigraphy, or NICE Guidelines on Subsequent Unnecessary Angiography Rates: The CE-MARC 2 Randomized Clinical Trial. JAMA. 2016;316:1051–1060.

39. Curzen N, Nicholas Z, Stuart B, Wilding S, Hill K, Shambrook J, Eminton Z, Ball D, Barrett C, Johnson L, Nuttall J, Fox K, Connolly D, O'Kane P, Hobson A, Chauhan A, Uren N, Mccann G, Berry C, Carter J, Roobottom C, Mamas M, Rajani R, Ford I, Douglas P, Hlatky M, Investigators on B of the F. Fractional flow reserve derived

from computed tomography coronary angiography in the assessment and management of stable chest pain: the FORECAST randomized trial. Eur Heart J [Internet]. 2021;Available from: https://doi.org/10.1093/eurheartj/ehab444

40. Maron DJ, Hochman JS, Reynolds HR, Bangalore S, O'Brien SM, Boden WE, Chaitman BR, Senior R, López-Sendón J, Alexander KP, Lopes RD, Shaw LJ, Berger JS, Newman JD, Sidhu MS, Goodman SG, Ruzyllo W, Gosselin G, Maggioni AP, White HD, Bhargava B, Min JK, Mancini GBJ, Berman DS, Picard MH, Kwong RY, Ali ZA, Mark DB, Spertus JA, Krishnan MN, Elghamaz A, Moorthy N, Hueb WA, Demkow M, Mavromatis K, Bockeria O, Peteiro J, Miller TD, Szwed H, Doerr R, Keltai M, Selvanayagam JB, Steg PG, Held C, Kohsaka S, Mavromichalis S, Kirby R, Jeffries NO, Harrell FE, Rockhold FW, Broderick S, Ferguson TB, Williams DO, Harrington RA, Stone GW, Rosenberg Y. Initial Invasive or Conservative Strategy for Stable Coronary Disease. N Engl J Med [Internet]. 2020;382:1395–1407. Available from: https://doi.org/10.1056/NEJMoa1915922

41. Altman DG, Bland JM. Missing data. BMJ. 2007;334:424.

42. Shinkins B, Thompson M, Mallett S, Perera R. Diagnostic accuracy studies: how to report and analyse inconclusive test results. BMJ Br Med J [Internet]. 2013;346:f2778. Available from: http://www.bmj.com/content/346/bmj.f2778.abstract

43. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Heal. 2020;2:e489–e492.

Table 1. Subject baseline characteristics for the datasets used to develop and validate the CNNs and predictive model. There is some overlap between the various datasets however there is a complete separate between any data used for training and any used for validation.

	Training dataset	**Testing dataset**
Number of Studies	578	154
Demographics		
Age (years)	64.5±11.5	61±10.7
Sex (% M)	54.3	50
BMI (kg/m^2)	29.9±6.2	30.1±7.9
Hypertension (%)	40.9	62.1
Type I Diabetes (%)	0.34	-
Type II Diabetes (%)	16.7	25.6
Previous MI (%)	7.1	7.8
Previous PCI (%)	30.8	8.4
CAD Family History (%)	14.7	14.9
Resting EF (%)	61.5±11.0	58.3±9.8
Angiography:		
Severe Stenosis (N)	85	45
Single Vessel Disease (%)	47.6	51.1
Two Vessel Disease (%)	36.9	37.8
Three Vessel Disease (%)	15.5	11.1
Protocols:		
Exercise Stress (%)	12.8	64.3
Contrast Enhancement (%)	69.4	98.7
Machine:		
GE	184	6
Vivid 7	32	0
Vivid E9	150	6
Vivid E95	2	0
Phillips	394	148
iE33	265	147
EPIQ 7G	2	0
EPIQ 7C	127	1

Figure Legends

Figure 1. Regional wall motion abnormalities identified at peak stress
Mean peak stress regional wall motion abnormalities as scored by trained echocardiographers from 0-5 where 0 is normal wall motion (left hand panel) compared to a concentration of relevant features for each segment identified by the AI model as predictive of clinical outcome (right hand panel).

Figure 2. Disease stratification and classification capabilities.
Panel A presents example ROC curves of three selected features. Feature 1 = A2C at stress rectangularity feature 1, Feature 2 = A4C at stress velocity feature 14, Feature 3 = SAX at rest elliptical variance feature 1. Panels B, C & D show plots of feature values with individuals who had positive clinical outcome coloured orange and negative outcomes coloured blue to demonstrates capability of individual model features to differentiate outcome. Vertical and horizontal lines indicate example cut-off values for disease classification, optimised for balanced sensitivity and specificity. Panel E demonstrates performance of the AI-based classifier on training and independent validation datasets based on ROC curves. Disease stratification is based on an ensemble model built from 31 of the novel features. Feature nomenclature for panels B, C & D: Feature names begin with an indication of the view and stage of the SE examination (4P = A4C view at stress, SAX_R = parasternal short axis view at rest), followed by the measurement (e.g. rectangularity, tortuosity, velocity).

Figure 3. Evaluation of reader study.
Panel A is a ROC curve of stress echocardiography interpretation of four readers with (solid line) and without (dashed line) the assistance of the AI-based classification. Panel B shows the inter-reader agreement for positive (positive agreement) and negative (negative agreement) clinical interpretations for unassisted and AI-assisted reads**.** Panel C shows individual reader performance with (solid) and without (dashed) the assistance of the AI classifier.

Central Illustration. Novel artificial intelligence derived features improve coronary disease classification
Novel quantitative features of regional wall motion can be implemented into machine learning classifiers to assist and enhance clinician performance during interpretation of stress echocardiography in the investigation of coronary artery disease.