openheart End-to-end deep-learning model for the detection of coronary artery stenosis on coronary CT images

Vibha Gupta ^(b), ¹ Petur Petursson, ² Aidin Rawshani, ¹ Jan Boren, ^{1,3} Truls Ramunddal, ² Deepak L Bhatt ^(b), ⁴ Elmir Omerovic ^(b), ^{1,2} Oskar Angerås, ² Gustav Smith, ^{1,2,3} Naveed Sattar, ⁵ Erik Andersson ^(b), ¹ Björn Redfors, ^{2,3} Lukas Hilgendorf ^(b), ¹ Göran Bergström, ¹ Carlo Pirazzi, ² Kristofer Skoglund ^(b), ^{1,2} Araz Rawshani^{1,2}

Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/ openhrt-2024-002998).

To cite: Gupta V, Petursson P, Rawshani A, et al. End-to-end deep-learning model for the detection of coronary artery stenosis on coronary CT images. *Open Heart* 2025;**12**:e002998. doi:10.1136/ openhrt-2024-002998

Received 4 October 2024 Accepted 26 December 2024

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Dr Vibha Gupta; gupta85vibha19085@gmail.com

ABSTRACT

Purpose We examined whether end-to-end deep-learning models could detect moderate (\geq 50%) or severe (\geq 70%) stenosis in the left anterior descending artery (LAD), right coronary artery (RCA) or left circumflex artery (LCX) in iodine contrast-enhanced ECG-gated coronary CT angiography (CCTA) scans.

Methods From a database of 6293 CCTA scans, we used pre-existing curved multiplanar reformations (CMR) images of the LAD, RCA and LCX arteries to create end-to-end deep-learning models for the detection of moderate or severe stenoses. We preprocessed the images by exploiting domain knowledge and employed a transfer learning approach using EfficientNet, ResNet, DenseNet and Inception-ResNet, with a class-weighted strategy optimised through cross-validation. Heatmaps were generated to indicate critical areas identified by the models, aiding clinicians in understanding the model's decision-making process.

Results Among the 900 CMR cases, 279 involved the LAD artery, 259 the RCA artery and 253 the LCX artery. EfficientNet models outperformed others, with EfficientNetB3 and EfficientNetB0 demonstrating the highest accuracy for LAD, EfficientNetB2 for RCA and EfficientNetB0 for LCX. The area under the curve for receiver operating characteristic (AUROC) reached 0.95 for moderate and 0.94 for severe stenosis in the LAD. For the RCA, the AUROC was 0.92 for both moderate and severe stenosis detection. The LCX achieved an AUROC of 0.88 for the detection of moderate stenoses, though the calibration curve exhibited significant overestimation. Calibration curves matched probabilities for the LAD but showed discrepancies for the RCA. Heatmap visualisations confirmed the models' precision in delineating stenotic lesions. Decision curve analysis and net reclassification index assessments reinforced the efficacy of EfficientNet models, confirming their superior diagnostic capabilities.

Conclusion Our end-to-end deep-learning model demonstrates, for the LAD artery, excellent discriminatory ability and calibration during internal validation, despite a small dataset used to train the network. The model reliably produces precise, highly interpretable images.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previous research has demonstrated the potential of deep-learning models for detecting coronary artery stenoses in curved multiplanar reformations (CMR) images. However, most studies have relied on consistent, standardised data from controlled protocols, used relatively small datasets and reported good performance at various levels.

WHAT THIS STUDY ADDS

⇒ This study uses an end-to-end deep-learning approach trained on routine clinical CMR images to detect moderate and severe stenoses in the left anterior descending artery (LAD), right coronary artery (RCA) and left circumflex artery (LCX). It addresses anatomical challenges with diverse, real-world data from multiple hospitals. Advanced preprocessing techniques and interpretable heatmaps enhance model performance and promote clinical adoption.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Integrating routine clinical data into model training enhances diagnostic accuracy and promotes the adoption of deep-learning models in clinical workflows. With area under the curve for receiver operating characteristic up to 0.95 for the LAD, reasonable performance for the RCA and LCX, well-calibrated predictions and interpretable heatmaps, these findings support efficient management of coronary artery disease, improving patient care and guiding future research.

INTRODUCTION

Coronary artery disease is the leading cause of mortality worldwide.¹ Advances in coronary CT angiography (CCTA) have markedly enhanced the detection of atherosclerotic plaques within the coronary arteries.²³ Nextgeneration photon-counting CT scanners will further manifest the role of CCTA as the firstline tool for coronary imaging.⁴ CCTA allows for a comprehensive evaluation of stenoses,





1

STRENGTHS AND LIMITATIONS OF THIS STUDY

- \Rightarrow Achieved an area under the curve of 0.95 for moderate and 0.94 for severe stenosis in the left anterior descending artery (LAD). Calibration curves closely matched observed probabilities, indicating robust model performance.
- ⇒ EfficientNet models, optimised with transfer learning and a classweighted approach, surpass other deep-learning models in accuracy and performance in detecting stenosis in LAD and right coronary artery arteries.
- \Rightarrow Heatmaps provide visual confirmation of critical areas, enhancing the interpretability of the model's decision-making process.
- $\Rightarrow\,$ The study uses a very diverse dataset.
- $\Rightarrow\,$ The study is limited by a relatively small dataset.
- ⇒ The study has few positive cases, which may affect the robustness of the findings. Hence, for left circumflex artery, detection of severe (\geq 70%) stenosis could not be performed.
- \Rightarrow External validation was not conducted, which may limit the generalisability of the findings.
- ⇒ Potential inclusion bias may arise from the selection of curved multiplanar reformation images by radiologists, based on their clinical needs.

plaque composition, pericoronary tissues and other emerging predictors of coronary events.⁵ ⁶ The interpretation of CCTA images is time-consuming, resourceintensive and the human eye may overlook subtle nuances carrying significant clinical relevance. Deep learning has shown exceptional capabilities in cardiovascular diagnostics, surpassing human performance in several instances,^{7–10} including CCTA.¹¹

End-to-end deep-learning models offer the possibility to use raw, unprocessed and data. This, in conjunction with high-quality labelling, is a promising approach for CCTA and cardiovascular imaging in general.¹³ We explored the feasibility of developing end-to-end models using curved multiplanar reformations (CMR) images, thereby eliminating the need for complex segmentation procedures and manual interventions. Using various deep-learning architectures, we aimed to develop a fully automated model capable of detecting significant stenoses in the left anterior descending coronary artery (LAD), right coronary artery (RCA) and left circumflex artery (LCX).

METHODS

Patient selection

The original database for our study encompassed 6293 CCTA studies conducted from 2010 to 2021 in Västra Götaland County, Sweden, which serves a population of 1.6 million inhabitants. This database included patients of all ages and sexes who underwent iodine contrastenhanced ECG-gated CCTA as part of their medical management. Only accredited radiologists interpreted the CT studies, which were performed in accordance with guidelines regarding scanners, administration of beta-blockers and sublingual glyceryl nitrate. Most of these cases were referred primarily for the evaluation of the coronary arteries. To ensure the final model's applicability across a diverse patient spectrum, we did not exclude any studies based on the referral cause. Thus, the inclusion criterion for our study was the completion of an iodine contrast-enhanced ECG-gated CCTA, with an expert evaluation of the coronary arteries.

Online supplemental table 1 illustrates the detailed breakdown of the total cases for stenosis prevalence in the LAD, RCA and LCX artery datasets. The data distribution for training, testing and validation across both designed tasks,; (50% or greater occlusion) and (70% or greater occlusion), is also presented in online supplemental table 1. Due to the low prevalence of positive stenosis cases, we were unable to evaluate the LCX for the detection of 70% or greater stenoses.

Curved multiplanar reformations

We included all generated, by the radiologist, CMR images. No new CMR images were created specifically for this study. Overall, 185 500 CMR images were generated across roughly 900 patients, representing 14% of the individuals in the original database. The number of images per reconstruction varied, with counts ranging from 6 to 128 images per CMR case. While it is true that the use of CMR images, generated based on the radiologists' discretion, could introduce a selection bias, the decisions were guided by the specific clinical needs and criticality of each case. For the main analysis, we focused on studies that included a standard stack of 36 images per CMR, which was the most common configuration.

In a supplementary analysis, we constructed models that processed a varying number of CMR images, ranging from 18 to 128, to predict moderate or greater stenosis in the LAD and RCA. This experiment aimed to assess the impact of variable image counts on model performance, reflecting a probable real-world application scenario.

Label extraction

The labels guiding the supervised model were retrieved from the radiologist's final evaluation of the scan. The report included information regarding stenoses, plaque types, calcium scores and other clinically relevant findings. The written report was used to determine the level of stenosis in coronary artery segments 1-15. One physician trained in CCTA (level 1) interpreted the report to classify each segment, and an interventional cardiologist provided second opinions in ambiguous cases. The reports used a graded classification of the degree of stenosis. The following cut-offs were used: 0% stenosis classified as no visible stenosis; 1%-24% as minimal stenosis; 25%-49% as mild stenosis; 50%-69% as moderate stenosis; 70%-99% as severe stenosis and 100% representing total occlusion. We did not assess studies that did not evaluate the coronary arteries (n=10).

The labels used for modelling were assigned to the entire artery rather than individual segments. The labels corresponded to the highest level of stenosis across the entire artery. We evaluated performance based on these artery-level assessments.

Stenosis severity task

We implemented a binary classification approach to categorise the severity of stenosis into two separate tasks. The first task was designed to identify cases with at least moderate (50% or greater) stenoses, while the second focused on detecting stenoses that were at least severe (70% or greater). This aimed to assess the model's capability to accurately predict different levels of narrowing.

Preprocessing

CMR images were extracted from DICOM files and cropped to a specific Hounsfield unit (HU), range (-1000 to 2000) to ensure consistent greyscale representation and remove outliers. Following this, the cropped images were normalised to a bitmap between 0 and 1. To eliminate non-cardiac areas, a mask was generated from the thresholded DICOM image, and a windowing range of 55-800 HU was applied to select cardiac tissue and intravenous contrast. All selected areas were converted to a binary bitmap, where voxels within the selected attenuation were set to true. This process resulted in multiple islands of selected voxels, with islands under 30 voxels being removed. Given that the previous windowing primarily selected intravenous contrast and some cardiac tissue, a 10-voxel wide border was added to the mask to ensure the capture of all soft plaque and tissue surrounding the vessels. Consequently, a mask comprising multiple islands was obtained, from which the largest island was selected and combined with the normalised bitmap image to produce the final cropped bitmap. Figure 1 illustrates the flow of the preprocessing pipeline and provides a visual representation of each step.

Model building and evaluation

The entire pipeline, including data fetching, preprocessing, model training, data splitting and evaluation, is illustrated comprehensively in online supplemental figure 1.

We explored a range of models, including various EfficientNet¹⁴ models from E0 to E3, as well as ResNet¹⁵ and



unit.

DenseNet¹⁶ variations, and Inception-ResNet.¹⁷ Each model was fine-tuned to detect stenosis in these arteries. Following a transfer learning paradigm, we froze the initial layers and fine-tuned the models on our dataset. To ensure compatibility with our target image size, the 'top' parameter was set to false. We conducted experiments to optimise model performance by varying dropout rates from 0 to 0.5 and regularisation parameter (lambda) values from 0 to 0.001.

To mitigate the effects of class imbalance and prevent potential biases in the model, we implemented a class dictionary approach. This approach involved adjusting the class weights during training to give more importance to the minority class (positive samples) and less importance to the majority class (negative samples). This aimed to address overfitting and improve the model's ability to learn from both classes equally. Additionally, we used data augmentation techniques such as random rotation, horizontal and vertical shifts, zooming, and horizontal flipping. These augmentations enhanced the model's robustness and generalisation, contributing to its improved performance.

We used fivefold cross validation, repeated five times, with data split into training (70%), validation (10%) and testing (20%). Multiple metrics were used to assess model performance. These metrics included image recognition rate (IRR) which is overall accuracy, patient recognition rate (PRR), F1-score, precision (positive predictive value), specificity, and negative predictive value, area under the receiver operating characteristic curve (AUC-ROC), AUC precision-recall (AUC-PR), net benefit (NB), decision curve analysis, calibration curves, and the Brier score. Additionally, advanced metrics such as the additive and absolute net reclassification index (NRI) were used to evaluate the models' improvement in risk prediction, enhancing understanding of their clinical utility. Since many of these metrics are straightforward and easy to understand, a concise explanation of PRR is provided below, as it is not as common as others:

The PRR is defined as the average of patient recognition (PR) where PR is ratio of correctly classified images to the total number of images per case. This can be expressed as: $PRR = \frac{\sum_{k=1}^{N} PR_k}{N}$, N is total test CMR cases.

Interpretable deep learning

Heatmaps were generated to highlight the areas of focus used by each model during the prediction process. These heatmaps allowed visualisation of which regions of the input images were deemed most relevant by the models, providing insights into decision-making processes.

RESULTS

We identified a total of 900 cases with CMR images. For CMR cases containing 36 images, there were 279 for the LAD, 259 cases for the RCA and 253 cases for the LCX. When considering CMR cases with 18 or more images, the total number of CMRs was 640 for the LAD and 568 for the RCA.

The EfficientNet models consistently outperformed the other deep-learning models across all; LAD, RCA and LCX. Specifically, EfficientNetB3 exhibited robust performance for moderate or greater stenosis in LAD, while EfficientNetB0 emerged as the better model to detect severe or greater stenoses. For RCA, EfficientNetB2 demonstrated the highest performance for moderate and severe stenoses or greater. EfficientNetB0 demonstrated good performance in detecting moderate stenoses in the LCX. However, due to the very few positive cases of severe or greater stenoses in the LCX, we were unable to model the LCX for this task.

The results represent the best EfficientNet models, as judged by accuracy, AUC-ROC and F1 score. All reported metrics were evaluated on an artery basis, encompassing all segments corresponding to each artery.

Figure 2 illustrates AUC-ROC plots, CI and Brier scores for both the LAD and RCA. For moderate stenosis in the LAD (figure 2A), we achieved an AUC-ROC of 0.95 with a Brier score of 0.086 and a narrow 95% CI of 0.92 to 0.97. Similarly, severe stenosis in the LAD (figure 2B) exhibited an AUC-ROC of 0.94, a Brier score of 0.075 and a 95% CI of 0.89 to 0.99. In figure 2C, the model's performance in identifying moderate or greater stenosis in the RCA showed a mean AUC of 0.92 with a Brier score of 0.065 and a wider 95% CI of 0.85 to 0.95. Furthermore, figure 2D displays the performance for severe stenosis in the RCA, indicating a mean AUC of 0.92, a Brier score of 0.05 and a 95% CI of 0.87 to 0.97. The narrower CI for the LAD suggests more consistent model performance in predicting stenosis compared with the RCA, indicating greater reliability. This implies that the model's predictions for the LAD artery are more stable across different scenarios, enhancing confidence in clinical applications. However, this consistency is also evident through calibration and other metrics (later subsection).

Detailed AUC-ROC plots, depicting performance for each fold along with corresponding Brier scores, are presented in online supplemental figure 2. Each subplot illustrates six curves, representing fivefold and their mean, showcasing individual fold outcomes.

Along with AUC-ROC, AUC-PR was also calculated. The mean AUC-PR for LAD moderate or greater stenosis and severe or greater stenosis were 0.94 and 0.85, respectively, while for RCA, they were 0.79 and 0.56.

For moderate stenosis in the LCX, an AUC-ROC of 0.88 with a Brier score of 0.0692 and 95% CI of 0.78 to 0.98 was achieved.

Calibration

Figure 3 shows the calibration results for both prediction tasks for LAD and RCA. Figure 3A shows the calibration for the prediction of moderate stenoses in the LAD, whereas figure 3B shows the calibration for severe stenoses. In both panels, the calibration curve is well aligned with the observed probabilities, particularly below 50%

A LAD: Moderate or more stenosis

B LAD: Severe or more stenosis



Figure 2 The AUC-ROC along with 95% CI and Brier score for both LAD and RCA for all tasks. (A, B) LAD, (C, D) RCA. (A) Prediction of moderate or more stenosis in LAD. (B) Prediction of severe or more stenosis in LAD. (C) Prediction of moderate or more stenosis in RCA. (D) Prediction of severe or more stenosis in RCA. AUC-ROC, area under the curve for receiver operating characteristic; LAD, left anterior descending artery; RCA, right coronary artery.

probability (which may have the greatest impact on sensitivity).

Similarly, figure 3C,D presents calibration results for the RCA. Figure 3C displays calibration for the prediction of moderate stenoses and figure 3D shows the results for severe stenoses. The calibration curves in these figures highlight significant deviations from the ideal prediction, including both overestimation and underestimation of risk.

Figure 4 displays the evaluation performance for LCX moderate cases. It reveals that LCX achieved an AUROC of 0.88, accompanied by broader CIs (figure 4A,B). The calibration curve showed a significant overestimation of risk at higher (figure 4C). At thresholds ranging from 0% to 1%, sensitivity was observed between 0.8 and 0.78, with precision consistently below 60%. LCX performance was lower compared with LAD and RCA, likely due to fewer positive cases available for evaluation.

Threshold-dependent metrics

Threshold-dependent metrics were analysed across a decision threshold ranging from 0% to 50% probability of stenosis. Sensitivity, specificity, precision, negative predictive value, IRR and PRR were plotted against this threshold range in online supplemental figure 3.

In online supplemental figure 3A, depicting moderate stenosis in the LAD, it was observed that even at low thresholds (0.05 (5%) to 0.1 (10%)), sensitivity and negative predictive value consistently exceeded 90%, and the PRR was above 70%. Precision and specificity increased from 0.6 to 0.7, 0.6 to 0.78, respectively. Furthermore, with increasing decision thresholds up to 0.2, sensitivity and negative predictive value remained relatively stable in the 90s. Meanwhile, the PRR, specificity and precision showed rapid increases, approaching 85%. This suggests that adjusting the threshold within this range can

A LAD: Moderate or more stenosis

B LAD: Severe or more stenosis



Figure 3 The calibration curve for all cases and tasks. (A) Calibration results for predicting moderate stenosis in LAD. (B) Calibration results for severe stenosis in LAD. (C) Calibration results for predicting moderate stenosis in RCA. (D) Calibration results for severe stenosis in RCA. (A) Calibration results for severe stenosis in RCA. (D) Calibration results for severe steno

significantly impact these metrics, particularly precision and specificity, without substantially affecting sensitivity and negative predictive value.

In online supplemental figure 3B (severe stenosis in LAD), steeper positive changes in the range of 0.05–0.5 were noticed for precision, specificity and PRR, while the negative predictive value did not change much. Sensitivity dropped from 0.9 to 0.78 within the threshold range of 0.05–0.5.

From online supplemental figure 3C,D (representing moderate and severe stenosis in RCA), it is observed that sensitivity drops to 0.53 and 0.44, respectively, as the threshold reaches 50%. In comparison, precision reaches 0.82 and 0.68 at the same threshold. The PRR remains around 0.90 or higher for most thresholds.

Heatmaps for interpretable deep learning

Figure 5 illustrates a colour-coded heatmap generated by the best-performing model for both LAD and RCA test

cases. This figure includes four examples for each artery, where each example shows the original images alongside blended images. The blended images overlaid on the original CCTA image with colours indicating the areas of focus by the model. The side activation bar shows the value of each colour and its intensity in predicting the outcome. From the figure, it is evident that the model performs well in detecting calcified plaque but struggles to identify soft plaque, which has a very similar density to surrounding tissue. This indicates a limitation in the model's ability to differentiate between soft plaque and surrounding structures.

To compare the heatmaps, generated by a series of EfficientNet, we put additional figure in online supplemental appendix. Online supplemental figure 4 shows an example of a colour-coded heatmap, corresponding to the output from the three best-performing neural network architectures (EfficientNetB3, EfficientNetB2

в



Figure 4 The AUC-ROC along with 95% CI and Brier score, calibration, evaluation metrics (threshold range 1%–50%) for LCX. (A) AUROC curve for each fold along with mean. (B) Mean AUROC with 95% CI. (C) Calibration curve. (D) Valuation of various metrics on the threshold range from 0% to 50%. AUC-ROC, area under the curve for receiver operating characteristic; LCX, left circumflex artery.

and EfficientNetB0). These heatmaps represent areas where the neural networks are focusing to make a determination about the presence of stenosis. EfficientNetB3 shows a heatmap with a concentration of warm colours (reds and yellows) around the stenosis, indicating high attention in that region, suggesting that the model predicts this area as being significant for the diagnosis of stenosis. The EfficientNetB2 heatmap shows a similar pattern, with warm colours around the stenosis, but with a more diffused pattern extending along the artery and in non-relevant regions. Lastly, the EfficientNetB0 heatmap also highlights the area of stenosis, with a very concentrated area at the stenosis, with less spread around the lesion. Overall, the models exhibited greater attention to calcified plaques and were less attentive to soft plaques.

Δ

Models using variable number of CMR images

Online supplemental figures 5–8 show the performance of LAD and RCA for both tasks with varying numbers of CMR images per case. As evident in online supplemental figure 5, we achieved similar model performance for LAD when using a variable number of images per CMR. An AUC-ROC of 0.93 was achieved for LAD moderate or greater stenosis, with a narrow CI. Calibration showed slight underestimation in the lower probability range and overestimation in the higher probability range. This behaviour was also observed for LAD severe or greater stenosis (online supplemental figure 6) with an AUC-ROC of 0.92. For RCA (online supplemental figures 7,8), however, calibration appeared mostly overestimated, and the mean AUC-ROC dropped.

Decision analysis curve

This analysis (online supplemental figure 9) shows how the NB behaved across decision thresholds. Efficient-NetB3 for moderate stenosis exhibits high NB across the low threshold range (0–0.292), a critical range for detecting subtle stenoses. Thus, EfficientNetB3 was considered the best model for the task.

Online supplemental figure 10 shows corresponding results for predicting severe stenosis in LAD. It is noted that EfficientNetB0 shows higher performance in the critical threshold ranges.

Additive and absolute NRI

The NRI offers additional insight into how effectively a new model enhances risk prediction compared with a

А

В

Original Image

Original Image

Original Imag



Figure 5 Heatmaps for LAD and RCA. (A) Heatmaps of LAD test cases. (B) Heatmaps of RCA test cases. LAD, left anterior descending artery; RCA, right coronary artery.

reference model. The EfficientNetB3 model achieves the highest NRI index (both additive and absolute; online supplemental figure 11) for predicting moderate stenosis in LAD. Similarly, for severe stenosis in LAD, the EfficientNetB0 model exhibited favourable performance (online supplemental figure 12).

DISCUSSION

We developed an end-to-end model capable of detecting coronary artery stenoses with very high accuracy, despite a relatively small training set. This underscores the potential for end-to-end deep-learning models to attain clinically meaningful results even with limited training data. We expect that model performance will increase substantially with the use of all 6322 cases in our dataset. Indeed, the development of an accurate deep-learning model for CCTA analysis could be a game-changer for outpatient investigations and emergency room settings, where swift and precise triage of chest pain patients is crucial. Immediate and reliable detection of atherosclerotic plaques and acute coronary syndromes will improve clinical decision-making, save lives and reduce healthcare costs.

Other research groups have also used CMR images to detect coronary stenoses. To the best of our knowledge, the largest study to date is the CNN-CASS study (n=828), which deployed a Shuffle Net V2-based approach, with a fixed 50-image stack per CMR, achieving 80% patientlevel accuracy for stenosis classification.¹⁸ Employing a token-mixer architecture (ConvMixer), Penso et al achieved high sensitivity for significant stenosis classification.¹⁹ While they did not report calibration plots, our model for LAD performs better. A transformer network with self-supervised learning by Bian et al obtained impressive accuracy and specificity using only 78 patients for training and testing. However, our study differs by not only providing high accuracy but also delivering interpretability through heatmaps and broader applicability by not restricting image counts per CMR. Unfortunately, the omission of calibration plots in most of the previously published studies make the comparison with our models difficult.²⁰ Zreik *et al* used a multitask recurrent convolutional neural network (RCNN) on CMR images from 163 scans. The RCNN performed two simultaneous classification tasks for plaque type (morphology) and stenosis severity. Assessing the detection of significant stenosis versus no or non-significant stenosis, the method attained accuracies of 0.94, 0.93 and 0.85 at the segment, artery and patient levels, respectively.²¹ Hampe *et al* used CCTA and deep learning to predicting the functional significance of stenosis using invasively measured FFR (fractional flow reserve) as labels.²² Although they report a moderate AUC, predicting functional significance is a critical task. We aim to pursue this task using our dataset which includes >1000 invasive FFR measurements.

It is important to recognise that our study design presented a considerable challenge for the neural networks. The task at hand was to identify the occurrence of either moderate (the first model) or severe (as per the second model) stenosis in any segment of the LAD or RCA or LCX, given the entire stack of CMR images as input, including with a variable number of input size. Thus, the impressive results, particularly for the LAD, demonstrate the opportunities for end-to-end models and transfer learning.

Although the models for the LAD artery achieved impressive results, the performance of the models for the RCA and LCX was less efficient. The relatively uniform anatomy of the RCA, which in a majority of cases gives rise to the posterior descending artery, suggests that anatomical complexity is an unlikely cause. Additionally, there is no evidence indicating an increased incidence of artefacts in RCA reconstructions. The most conspicuous factor is the limited size of the RCA training dataset, which almost certainly impacts model performance. The lower performance of the LCX can be attributed to the very small number of stenosis prevalent cases and its more complex anatomy compared with the LAD. Furthermore, class imbalance within the training set and a lower prevalence of stenoses could further hamper the neural network's ability to discern abnormalities. Finally, potential ambiguities in the RCA and LCX labelling process might contribute to the inferior model performance.

We deployed a comprehensive approach, which involved training various models with different architectures and optimisation strategies, ultimately resulting in AUC-ROC scores surpassing 94%. However, the tendency of these models to overestimate and underestimate risks necessitates a careful balance between discriminative power and calibration accuracy to refine risk prediction. Obviously, it is important to bear in mind that these results were derived using internal testing data. External validation is required to corroborate these findings. To support the replication of our study, we have made the preprocessing and training code available on GitHub (https://github.com/Vibha190685/DL-for-Detection-of-Coronary-Artery-Stenosis/tree/master).

Interpretable AI is key to bolster model transparency and build clinician trust. Therefore, we created heatmap visualisations explaining the decisions of the prediction model. These heatmaps pinpoint the areas that the models identify as critical for stenosis prediction within the CCTA images, thereby elucidating the rationale behind the models' decisions. This aims to boost trust in the application of deep learning within a clinical context. Although deep learning is likely to become the principal resource for clinical decision-making in the future, we anticipate a series of transitional stages. In these phases, a collaborative evaluation by human experts will be paramount, initially positioning human judgement at the forefront of diagnostic decisions.

Despite data limitations, our models demonstrated robust performance across multiple diagnostic tasks. EfficientNet series outperformed models like ResNet and DenseNet due to its balanced compound scaling method, scaling depth, width and resolution uniformly, ensuring better performance without significantly increasing model size. We believe that the preprocessing techniques employed refined the models' focus, effectively filtering out irrelevant data to concentrate on relevant features. As judged by the heatmaps (online supplemental material), our models showed a higher proficiency in identifying calcified plaques compared with low-attenuating (soft) plaques. Recognising soft plaques is critical as they are more susceptible to future acute coronary syndromes.^{23 24} We anticipate that expanding our dataset to include 18000 cases will enhance the model's ability to accurately detect high-risk plaques. This expansion enables us to conduct individual assessmentd of both soft and calcified plaque, further refining the model's performance in these critical domains. The addition of more data, including additional cases of RCA and LCX, will allow us to improve the model's performance for these specific cases.

In conclusion, our study marks a significant advance in the fusion of deep learning and cardiac imaging, providing a powerful predictive model that stands to benefit from an enlarged training dataset and further external validation. The model's end-to-end architecture is a cornerstone of its strength, supporting ongoing improvement and large-scale deployment. Moreover, its interpretability enhances its clinical value, bridging the gap between artificial intelligence and practical healthcare applications.

Author affiliations

¹Department of Molecular and Clinical Medicine, University of Gothenburg Institute of Medicine, Gothenburg, Sweden

²Department of Cardiology, Sahlgrenska universitetssjukhuset, Goteborg, Sweden ³Sahlgrenska Center for Cardiovascular and Metabolic Research, Institute of Medicine, Wallenberg Laboratory, Sahlgrenska University Hospital, Gothenburg, Sweden

⁴Icahn School of Medicine at Mount Sinai, New York, New York, USA

 $^{5}\mathrm{Institute}$ of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

Contributors VG and ArR designed the study. VG performed the calculations and drafted the first version of the manuscript. PP, ArR and LH provided critical validation of the results and expertise in interpretation related to modelling and preprocessing. All authors contributed to the interpretation of the data and the

revision of the manuscript. ArR is the supervising guarantor of the study. Al has been used to assist in and accelerate coding and debugging Python code for model creation, all of the generated code and suggestions have been checked for functionality and correctness by the corresponding author.

Funding This study is supported by a generous donation from the Knut and Alice Wallenberg Foundation and funding from the University of Gothenburg and the Region Västra Götaland to the Wallenberg Centre for Molecular and Translational Medicine (WCMTM) in Gothenburg.

Competing interests Dr. Bhatt discloses the following relationships: - Advisory Board: Angiowave, Bayer, Boehringer Ingelheim, CellProthera, Cereno Scientific, E-Star Biotech, High Enroll, Janssen, Level Ex, McKinsev, Medscape Cardiology, Merck, NirvaMed, Novo Nordisk, Stasys; Tourmaline Bio; Board of Directors: American Heart Association New York City, Angiowave (stock options), Bristol Myers Squibb (stock), DRS.LINQ (stock options), High Enroll (stock); Consultant: Broadview Ventures, Corcept Therapeutics, GlaxoSmithKline, Hims, SFJ, Summa Therapeutics, Youngene: Data Monitoring Committees; Acesion Pharma, Assistance Publique-Hôpitaux de Paris, Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute, for the PORTICO trial, funded by St. Jude Medical, now Abbott), Boston Scientific (Chair, PEITHO trial), Cleveland Clinic, Contego Medical (Chair, PERFORMANCE 2), Duke Clinical Research Institute, Mayo Clinic, Mount Sinai School of Medicine (for the ENVISAGE trial, funded by Daiichi Sankyo; for the ABILITY-DM trial, funded by Concept Medical; for ALLAY-HF, funded by Alleviant Medical), Novartis, Population Health Research Institute; Rutgers University (for the NIH-funded MINT Trial); Honoraria: American College of Cardiology (Senior Associate Editor, Clinical Trials and News, ACC.org; Chair, ACC Accreditation Oversight Committee). Arnold and Porter law firm (work related to Sanofi/Bristol-Myers Squibb clopidogrel litigation), Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute; AEGIS-II executive committee funded by CSL Behring), Belvoir Publications (Editor in Chief, Harvard Heart Letter), Canadian Medical and Surgical Knowledge Translation Research Group (clinical trial steering committees), CSL Behring (AHA lecture), Cowen and Company, Duke Clinical Research Institute (clinical trial steering committees, including for the PRONOUNCE trial, funded by Ferring Pharmaceuticals), HMP Global (Editor in Chief, Journal of Invasive Cardiology), Journal of the American College of Cardiology (Guest Editor; Associate Editor), Level Ex. Medtelligence/ReachMD (CME steering committees). MJH Life Sciences, Oakstone CME (Course Director, Comprehensive Review of Interventional Cardiology), Piper Sandler, Population Health Research Institute (for the COMPASS operations committee, publications committee, steering committee, and USA national co-leader, funded by Bayer), WebMD (CME steering committees), Wiley (steering committee); Other: Clinical Cardiology (Deputy Editor); Patent: Sotagliflozin (named on a patent for sotagliflozin assigned to Brigham and Women's Hospital who assigned to Lexicon; neither I nor Brigham and Women's Hospital receive any income from this patent); Research Funding: Abbott, Acesion Pharma, Afimmune, Aker Biomarine, Alnylam, Amarin, Amgen, AstraZeneca, Bayer, Beren, Boehringer Ingelheim, Boston Scientific, Bristol-Myers Squibb, Cardax, CellProthera, Cereno Scientific, Chiesi, CinCor, Cleerly, CSL Behring, Faraday Pharmaceuticals, Ferring Pharmaceuticals, Fractyl, Garmin, HLS Therapeutics, Idorsia, Ironwood, Ischemix, Janssen, Javelin, Lexicon, Lilly, Medtronic, Merck, Moderna, MyoKardia, NirvaMed, Novartis, Novo Nordisk, Otsuka, Owkin, Pfizer, PhaseBio, PLx Pharma, Recardio, Regeneron, Reid Hoffman Foundation, Roche, Sanofi, Stasys, Synaptic, The Medicines Company, Youngene, 89Bio; Royalties: Elsevier (Editor, Braunwald's Heart Disease); Site Co-Investigator: Cleerly.

Patient consent for publication Not applicable.

Ethics approval The study was approved by Swedish Ethical Review Authority (No. 2023-03311-02)

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Data cannot be shared/public due to ethical and privacy considerations. It includes sensitive patient information and disease diagnoses specific to the hospital setting.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

Correction notice This paper has been corrected after it was published online. Disclosure for Dr Bhatt has been added in the competing interest section of the paper, ORCID number was added and affiliation corrected.

ORCID iDs

Vibha Gupta http://orcid.org/0000-0002-3785-8380 Deepak L Bhatt http://orcid.org/0000-0002-1278-6245 Elmir Omerovic http://orcid.org/0000-0002-3875-8621 Erik Andersson http://orcid.org/0009-0001-0567-2020 Lukas Hilgendorf http://orcid.org/0009-0000-9518-3644 Kristofer Skoglund http://orcid.org/0000-0002-9095-1218

REFERENCES

- 1 Tsao CW, Aday AW, Almarzooq ZI, *et al*. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. *Circulation* 2022;145:e153–639.
- 2 SCOT-HEART Investigators, Newby DE, Adamson PD, et al. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. N Engl J Med 2018;379:924–33.
- 3 Hoffmann U, Ferencik M, Cury RC, et al. Coronary CT angiography. J Nucl Med Off Publ Soc Nucl Med 2006;47:797–806.
- 4 Si-Mohamed SA, Boccalini S, Lacombe H, et al. Coronary CT Angiography with Photon-counting CT: First-In-Human Results. *Radiology* 2022;303:303–13.
- 5 Sagris M, Antonopoulos AS, Simantiris S, et al. Pericoronary fat attenuation index-a new imaging biomarker and its diagnostic and prognostic utility: a systematic review and meta-analysis. Eur Heart J Cardiovasc Imaging 2022;23:e526–36.
- 6 Nurmohamed NS, van Rosendael AR, Danad I, et al. Atherosclerosis evaluation and cardiovascular risk estimation using coronary computed tomography angiography. Eur Heart J 2024;45:1783–800.
- 7 Zhang J, GajjalaS, Agrawal P, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation* 2018;138.
- 8 Al-Zaiti SS, Martin-Gill C, Zègre-Hemsey JK, et al. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. *Nat Med* 2023;29:1804–13.
- 9 Avram R, et al. CathAl: Fully Automated Interpretation of Coronary Angiograms Using Neural Networks, 2021. Available: https://doi.org/ 10.48550/arXiv.2106.07708
- 10 Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet* 2019;394:861–7.
- 11 Williams MC, Kwiecinski J, Doris M, et al. Low-Attenuation Noncalcified Plaque on Coronary Computed Tomography Angiography Predicts Myocardial Infarction: Results From the Multicenter SCOT-HEART Trial (Scottish Computed Tomography of the HEART). Circulation 2020;141:1452–62.
- 12 Lin A, Marral N, McElhinney P, *et al.* Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *Lancet Digit Health* 2022;4:e256–65.
- 13 Baskaran L, Maliakal G, Al'Aref SJ, *et al.* Identification and Quantification of Cardiovascular Structures From CCTA: An End-to-End, Rapid, Pixel-Wise, Deep-Learning Method. *JACC Cardiovasc Imaging* 2020;13:1163–71.
- 14 Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. 2019.
- 15 He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition, 2015. Available: https://doi.org/10.48550/arXiv.1512. 03385
- 16 Huang G, Liu Z, Maaten L, et al. Densely Connected Convolutional Networks, 2018. Available: https://doi.org/10.48550/arXiv.1608. 06993
- 17 Szegedy C, loffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI 2017;31.
- 18 Dobko M, Petryshak B, Dobosevych O. CNN-CASS: CNN for Classification of Coronary Artery Stenosis Score in MPR Images; 2020

- 19 Penso M, Moccia S, Caiani EG, et al. A token-mixer architecture for CAD-RADS classification of coronary stenosis on multiplanar reconstruction CT images. Comput Biol Med 2023;153:106484.
- 20 Bian Y, Ai D, Han T, *et al.* Transformer network with self-supervised learning for stenosis detection in ct angiography. ICAIP 2022; Zhanjiang China, November 18, 2022 10.1145/3577117.3577147 Available: https://dl.acm.org/doi/proceedings/10.1145/3577117
- 21 Zreik M, van Hamersvelt RW, Wolterink JM, et al. A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque

and Stenosis in Coronary CT Angiography. *IEEE Trans Med Imaging* 2019;38:1588–98.

- 22 Hampe N, van Velzen SGM, Planken RN, et al. Deep learningbased detection of functionally significant stenosis in coronary CT angiography. Front Cardiovasc Med 2022;9:964355.
- 23 Bhatt DL, Lopes RD, Harrington RA. Diagnosis and Treatment of Acute Coronary Syndromes: A Review. JAMA 2022;327:662–75.
- 24 Naghavi M, Libby P, Falk E, *et al*. From Vulnerable Plaque to Vulnerable Patient. *Circulation* 2003;108:1664–72.