SUPPLEMENTAL DATA

Clinical Data:

Past medical history and family history were prospectively collected in the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database. Follow up for MACE in the external testing population was obtained through the Discharge Abstracts/National Ambulatory Care Reporting system and Alberta Vital Statistics. MACE was defined as revascularization, non-fatal MI, admission for unstable angina, or all-cause mortality.

Myocardial Perfusion Image Analysis

Quality control for MPI was performed by experienced core laboratory technologists without knowledge of the clinical data. Stress and rest images were analyzed by Quantitative Perfusion SPECT (QPS) software (Cedars-Sinai Medical Center, Los Angeles, CA) as previously described to quantify total perfusion deficit (TPD) *(16).* TPD is a continuous measure which incorporates both the extent and severity of perfusion defects *(16).* Attenuation-corrected (AC) TPD was used for all analyses. Left ventricular ejection fraction (LVEF) was calculated from post-stress gated images.

CTAC Image Acquisition and Interpretation

At the University of Calgary, CTAC was performed using a built in CT scanner (Lightspeed VCT 64, GE, Boston, USA). CTAC imaging was performed after the rest acquisition during end-expiratory breath hold with no ECG-gating, in helical mode with a slice thickness of 5-mm, tube voltage of 120 kVp and 20 mA, using a 512x512 matrix. CTAC images were reviewed at the time of SPECT/CT MPI reporting and coronary calcium was graded visually as: absent, equivocal, present or extensive. Extensive calcification was defined as visually estimated CAC greater than 400. For comparisons with expert and DL annotated CAC scores, equivocal was combined with

present due to the low number of patients with equivocal visual CAC (n=28). Importantly, the expert visual estimates were informed by clinical information and perfusion findings. Details for the training population are available in Supplemental Table 1.

Expert CAC Annotations

The Expert reader annotation process included pixel by pixel calcification assignment into coronary calcification or non-coronary calcification using Cardiac Suite (Cedars Sinai Medical Center, Los Angeles, CA). Coronary calcium was annotated according to the involved vessel as LAD, LCX, LM and RCA. Non coronary calcification included calcification in the mitral valve, ascending aorta, descending aorta, aortic arch, aortic valve, tricuspid valve, pulmonary valve and pericardium. The DL model used these 2 categories to distinguish between coronary and non-coronary calcifications. CAC was quantified as previously described using the weighted sum of lesions with a density above 130 Hounsfield units, and multiplying the area of calcium by a factor related attenuation(*6*).

Model architecture

The model was built using PyTorch. We automatically segmented CAC from CTAC using a cascaded system of convLSTM(*17*). This system consists of two networks, first of which is trained for segmentation of the heart silhouette and the second network was trained to segment the CAC. The heart convLSTM was trained on a subset of training data with expert reader annotations for QFAT software(*18*). A supervised learning regime was used for both segmentation networks.

The input to the network consists of a CT slice(single slice input) along with the previous and next slices (Sequential input). This process was completely automated and there were no exclusions. The convolutional LSTM block takes in the sequential input to imitate the radiologist approach of sliding across various slices after looking at a single slice of interest. The output of the network

consists of an attention weighted combination of the results from the sequential input and the single slice of interest *(17)*. Softmax function is applied on the output to classify each pixel as background, coronary calcification or non-coronary calcification.

The heart mask was applied to the final CAC prediction to reduce any spurious bone overcalling or calcification in non-cardiac regions. In order to imitate the radiologist approach of aggregating information from adjacent slices, multiple slices were provided to both the networks as input and an attention map was generated by the convLSTM. The segmentation uses the attention weighted combination of the results from the sequential input and the single slice of interest which is later passed on to the softmax layer for final lesion mask creation. To counter the large class imbalance between CAC and background, we used subset sampling of the majority class as well as focal loss*(19)* as cost function between the ground truth expert reader annotation and network generated mask. The network was shown previously to have significantly reduced memory consumption for training and almost 2x faster inference times on a typical CPU*(17)*.

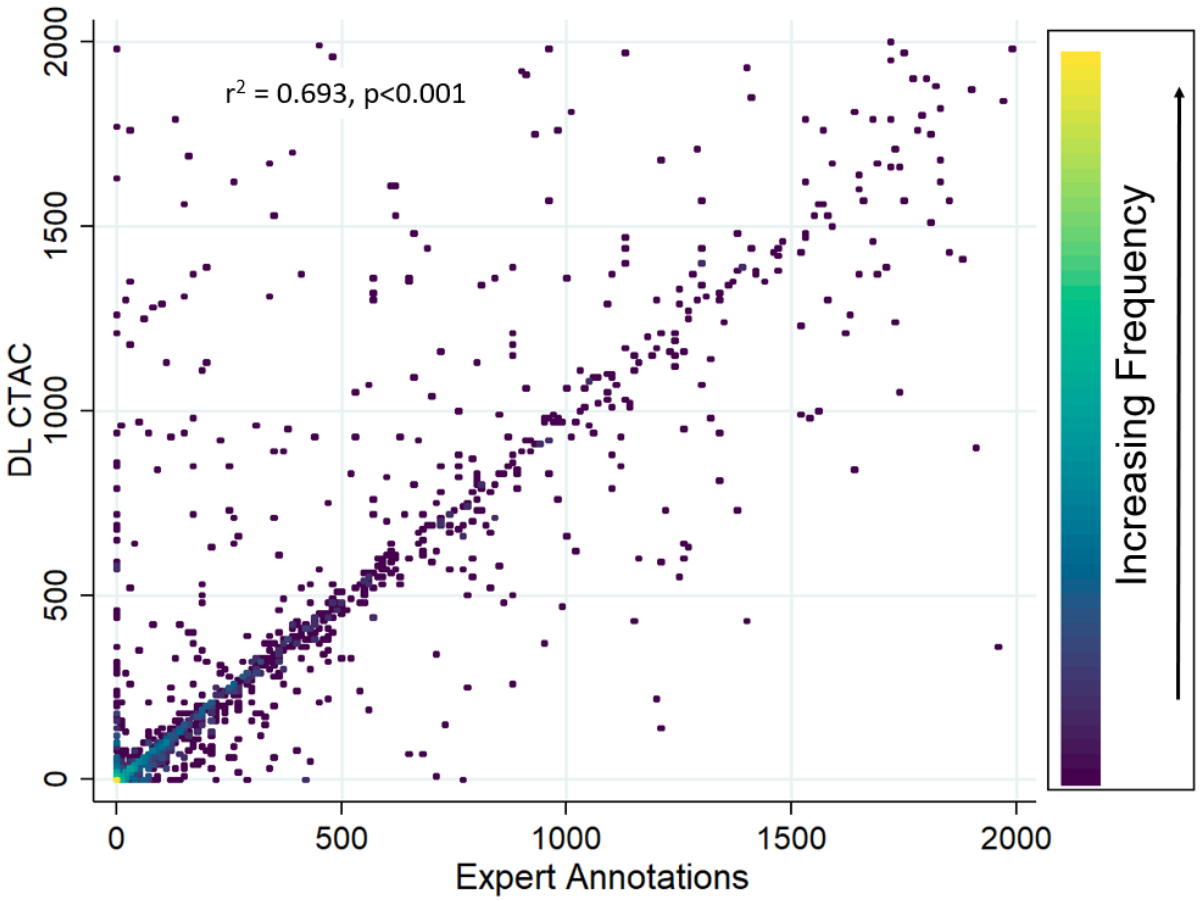
Statistical Analysis

Continuous variables were summarized as mean (standard deviation [SD]) if normally distributed and compared using a Student's t-test. Continuous variables that were not normally distributed were summarized as median (interquartile range [IQR]) and compared using a Mann-Whitney U-test. Associations with MACE were assessed with univariable and multivariable Cox proportional hazards analyses. Net reclassification index (NRI) was used to assess the additive prognostic utility of DL and expert annotated CAC. NRI was calculated when added to all other components of the multivariable model including: age, sex, past medical history, stress and rest AC TPD, and LVEF. Improvement in likelihood ratio chi-square (as a measure of model fit) and improvement in area under the receiver operating characteristic curve were also assessed. We also performed a

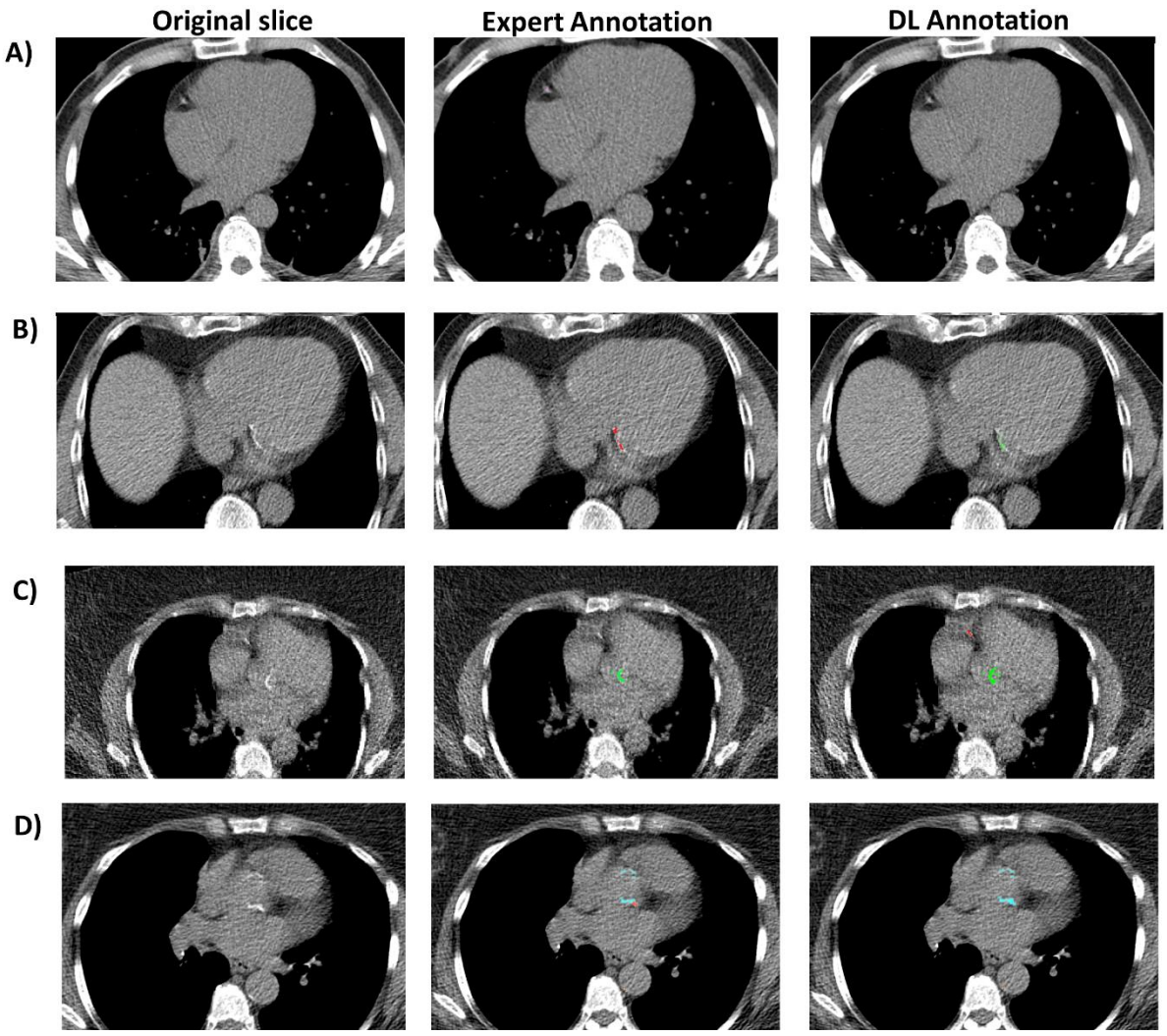
sensitivity analysis evaluating associations with hard adverse outcomes (death or MI) as well as an analysis in which patients who underwent early revascularization (revascularization within 90 days of SPECT/CT MPI) were excluded (n=52) since this may alter long-term outcomes (*20*,*21*). All analyses were performed using Stata/IC version 13.1 (StataCorp, College Station, Texas, USA) and R (version 4.1.2).

Review of Discrepant Cases

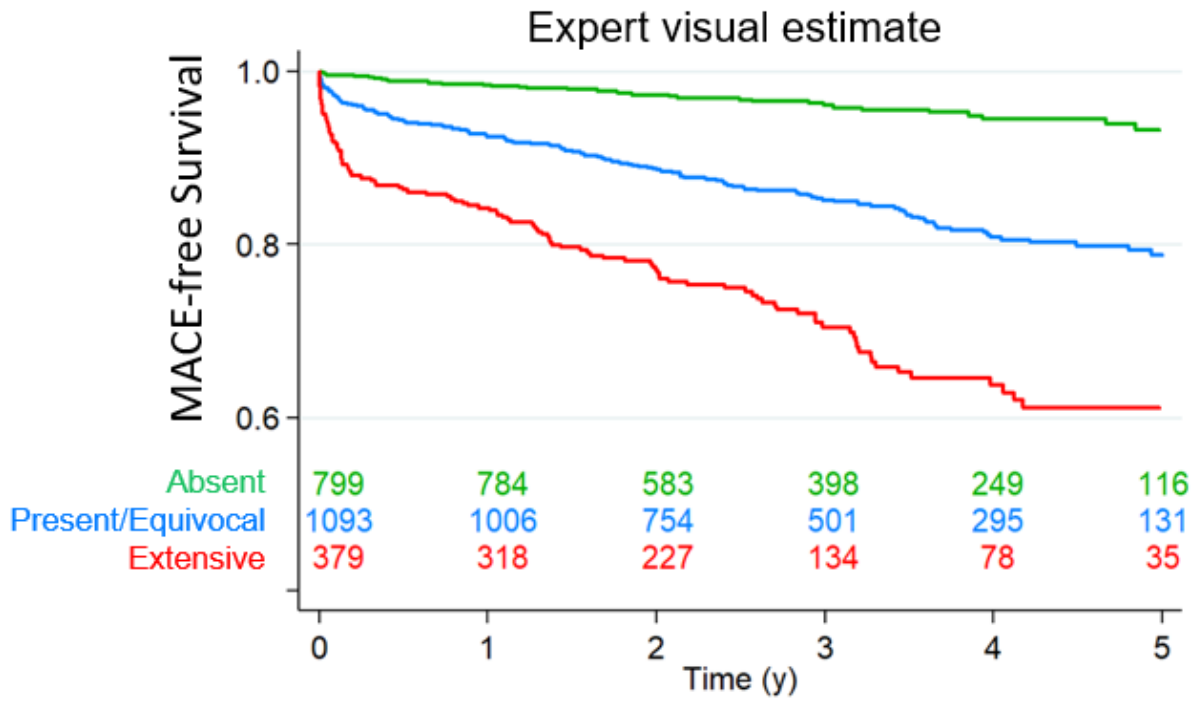
There were 142 cases with DL CAC of 0 with expert annotated CAC >0, which were reviewed manually to determine the most likely cause. This review identified image noise (n=125, 88%), confusion with valve-related calcium (n=13, 9%) or confusion with aortic calcification (n=10, 7%) as contributing to these cases. There were 196 cases with DL CAC >0 with expert annotated CAC of 0, which were reviewed manually to determine the most likely cause. This review identified image noise (n=159, 81%), confusion from pacing devices (n=19, 10%), confusion with valve-related calcium (n=9, 5%), confusion with aortic calcification (n=9, 5%), one case of calcified pericardium and one case of calcified lymph nodes as contributing to these cases.



Supplemental Figure 1: Pair-wise correlation between DL CTAC and expert annotations.



Supplemental Figure 2: Case examples of discrepant expert and deep learning (DL) annotations. In panel A, there is a small area of calcification identified by the expert in the right coronary artery (red) which was not identified by DL. In panel B, there is calcification that was annotated as belonging to the left circumflex by the expert reader (red) but attributed to mitral annular calcification by DL (green). In Panel C, there is aortic valve calcification noted by expert reader and DL (green), but also a small area of calcification noted in the RCA (red) which was annotated only by DL. In panel D, there is an area of calcification attributed to the left main artery by the expert reader (red) and the ascending aorta by DL (light blue).



Supplemental Figure 3. Kaplan-Meier survival curves for major adverse cardiovascular events (MACE). Increasing visual coronary artery calcium estimate was associated with increasing risk of MACE. Equivocal and present were considered as a single category due to the low number of patients with equivocal visually estimated coronary artery calcium (n=28).

	Training Population N = 6608	Site 1 n = 1827	Site 2 n = 4781	p-value	External Population N=2271
Age, Median (IQR)	63 (55, 72)	61 (53, 69)	64 (56, 73)	<0.001	69 (58, 74)
Male	3447 (52%)	796 (44%)	2651 (55%)	<0.001	1147 (51%)
Hypertension	4397 (68%)	1266 (70%)	3131 (67%)	0.038	1286 (57%)
Diabetes	1757 (27%)	487 (27%)	1270 (27%)	0.7	533 (23%)
Previous PCI	550 (8.5%)	4 (0.2%)	546 (12%)	<0.001	0 (0%)
Previous CABG	273 (4.2%)	0 (0%)	273 (5.8%)	<0.001	0 (0%)
Expert CAC, median (IQR)	52 (0, 480)	13 (0, 179)	88 (0, 602)	<0.001	15 (0, 208)
Expert CAC Category				<0.001	
CAC = 0	2196 (33%)	666 (36%)	1530 (32%)		962 (42%)
CAC 1-100	1508 (23%)	585 (32%)	923 (19%)		548 (24%)
CAC 101-400	1096 (17%)	281 (15%)	815 (17%)		362 (16%)
CAC >400	1808 (27%)	295 (16%)	1513 (32%)		399 (18%)
Slice Thickness (mm)				<0.001	
2.5	4511 (68%)	0 (0%)	4511 (94%)		0 (0%)
3	1827 (28%)	1827 (100%)	0 (0%)		0 (0%)
5	270 (4.1%)	0 (0%)	270 (5.6%)		2271 (100%)
Kilovolt potential				<0.001	
110	1 (<0.1%)	1 (<0.1%)	0 (0%)		0 (0%)
120	6401 (97%)	1620 (89%)	4781 (100%)		2271 (100%)
130	201 (3.0%)	201 (11%)	0 (0%)		0 (0%)
Unknown	5	5	0		
Tube current, median (IQR)	60 (60, 246)	416 (331, 568)	60 (60, 60)	<0.001	20 (20, 20)

Supplemental Table 1: Population characteristics for the training populations. CABG – coronary artery bypass grafting, CAC – coronary artery calcium, IQR – interquartile range, PCI – percutaneous coronary intervention.

Section	Checklist item	Location in the manuscript
1	**Designing the study plan**	
1.1	Describe the need for the application of machine learning to the dataset	Page 3, par 2
1.2	Describe the objectives of the machine learning analysis	Page 4, par 1
1.3	Define the study plan	Pages 6 and 7
1.4	Describe the summary statistics of baseline data	Page 7 and 8, Table 1
1.5	Describe the overall steps of machine learning workflow	Figure 1 and page 6 and 7
2	**Data standardization, feature engineering, and learning**	
2.1	Describe how the data were processed in order to make it clean, uniform, and consistent	Figure 1 and page 6 and 7
2.2	Describe whether variables were normalized and if so, how this was done	N/A
2.3	Provide details on the fraction of missing values (if any) and imputation methods	No missing values
2.4	Perform and describe feature selection process	N/A
2.5	Identify and describe the process to handle outliers if any	N/A
2.6	Describe whether class imbalance existed, and which method was applied to deal with it	Page 6 and 7
3	**Selection of Machine Learning Model**	
3.1	Explicitly define the goal of the analysis e.g., regression, classification, clustering	Page 6 and 7
3.2	Identify the proper learning method used (e.g., supervised, reinforcement learning etc.) to address the problem	Page 6 and 7
3.3	Provide explicit details on the use of simpler, complex, or ensemble models	N/A
3.4	Provide the comparison of complex models against simpler models if possible	N/A
3.5	Define ensemble methods, if used	N/A
3.6	Provide details on whether the model is interpretable	Page 6 and 7

4	**Model Assessment**	
4.1	Provide a clear description of data used for training, validation, and testing	Figure 1, page 6
4.2	Describe how the model parameters were optimized (e.g., optimization technique, number of model parameters etc.)	N/A
5	**Model Evaluation**	
5.1	Provide the metric(s) used to evaluate the performance of the model	Pages 10-12
5.2	Define the prevalence of disease and the choice of the scoring rule used	Page 6 and 8
5.3	Report any methods used to balance the numbers of subjects in each class	Page 6 and 7
5.4	Discuss the risk associated to misclassification	Page 9 and 10
6	**Best Practices for Model Replicability**	
6.1	Consider sharing code or scripts on public repository with appropriate copyright protection steps for further development and non-commercial use	Page 4
6.2	Release data dictionary with appropriate explanation of the variables	N/A
6.3	Document version of all software and external libraries	Page 7
7	**Reporting limitations, biases and alternatives**	
7.1	Identify and report the relevant model assumptions and findings	Page 12
7.2	If well-performing models were tested on a hold-out validation dataset, detail the data of that validation set with the same rigor as that of the training dataset (see section 2 above)	N/A

Supplemental Table 2. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME) Checklist

	Deep Learning Coronary Artery Calcium Score			
Visual Estimate	0	1-100	101-400	>400
Absent	623 (27.4%)	145 (6.4%)	17 (0.7%)	14 (0.6%)
Equivocal/Present	281 (12.4%)	429 (18.9%)	260 (11.4%)	123 (5.4%)
Extensive	4 (0.2%)	22 (1%)	77 (3.4%)	276 (12.2%)
	Expert Coronary Artery Calcium Score			
Visual Estimate	0	1-100	101-400	>400
Absent	740 (32.6%)	53 (2.3%)	2 (0.1%)	4 (0.2%)
Equivocal/Present	219 (9.6%)	478 (21%)	279 (12.3%)	117 (5.2%)
Extensive	3 (0.1%)	17 (0.7%)	81 (3.6%)	278 (12.2%)

Supplemental Table 3: Classification by visually estimated coronary artery calcification compared to deep-learning or expert annotated coronary artery calcium score. Equivocal and present were considered as a single category due to the low number of patients with equivocal visually estimated coronary artery calcium (n=28).

	No MACE n=1951	MACE n=320	p-value
DL CAC, median (IQR)	11.1 (0, 151.0)	178.0 (25.4, 851.9)	<0.001
DL CAC Categories			
CAC <1	856 (43.9%)	52 (16.3%)	<0.001
CAC 1 – 100	524 (26.9%)	72 (22.5%)	<0.001
CAC 101 – 400	273 (14.0%)	81 (25.3%)	<0.001
CAC > 400	298 (15.3%)	115 (35.9%)	<0.001
Age, median (IQR)	66.1 (58.1, 73.6)	70.7 (61.4, 77.2)	<0.001
Male, n(%)	946 (48.5%)	201 (62.8%)	<0.001
BMI, median (IQR)	29.8 (25.5, 33.1)	29.1 (24.9, 32.5)	0.069
Past Medical History, n(%)			
Hypertension	1092 (56.0%)	194 (60.6%)	0.12
Diabetes	432 (22.1%)	101 (31.6%)	<0.001
Dyslipidemia	844 (43.3%)	161 (50.3%)	0.019
Family history	978 (50.1%)	140 (43.8%)	0.029
Smoking	120 (6.2%)	30 (9.4%)	0.025
Stress AC TPD, median (IQR)	2.4 (0.8, 5.2)	6.2 (2.5, 13.9)	<0.001
Stress AC TPD Category			
Stress AC TPD < 1%	555 (28.4%)	36 (11.3%)	<0.001
Stress AC TPD 1 - < 5%	889 (45.6%)	109 (34.1%)	<0.001
Stress AC TPD 5 - <10%	337 (17.3%)	65 (20.3%)	<0.001
Stress AC TPD ≥ 10%	170 (8.7%)	110 (34.4%)	<0.001
Rest AC TPD, median (IQR)	0 (0, 0.6)	0.4 (0, 3.0)	<0.001
Stress LVEF, median (IQR)	67 (59, 74)	61 (51, 71)	<0.001

Supplemental Table 4. Patient characteristics in patients who experienced major adverse cardiovascular events (MACE) compared to those who did not. AC – attenuation correction, BMI – body mass index, CAC – coronary artery calcium, DL – deep learning, IQR – interquartile range, LVEF – left ventricular ejection fraction, TPD – total perfusion deficit.

	Death or myocardial infarction		Death	
	Adjusted HR (95% CI)	p-value	Adjusted HR (95% CI)	p-value
DL Categories				
CAC <1	Reference	Reference	Reference	Reference
CAC 1 – 100	1.67 (1.09 – 2.55)	0.018	1.53 (0.95 – 2.45)	0.078
CAC 101 – 400	2.38 (1.53 – 3.71)	<0.001	2.22 (1.37 – 3.60)	0.001
CAC > 400	2.55 (1.64 – 3.97)	<0.001	2.30 (1.42 – 3.74)	0.001
Expert Categories				
CAC <1	Reference	Reference	Reference	Reference
CAC 1 – 100	1.52 (0.98 – 2.35)	0.059	1.43 (0.88 – 2.31)	0.147
CAC 101 – 400	2.19 (1.42 – 3.39)	<0.001	2.06 (1.28 – 3.30)	0.003
CAC > 400	2.55 (1.64 – 3.92)	<0.001	2.15 (1.33 – 3.47)	0.002

Supplemental Table 5: Associations between coronary artery calcium (CAC) categories and secondary clinical outcomes. CI – confidence interval, HR – hazard ratio.

Deep Learning CAC	Increase in LR chi-square	50.4
	Increase in AUC	0.028 (0.010 to 0.046)
	Event NRI (95% CI)	0.230 (0.142 to 0.314)
	Non-event NRI (95% CI)	0.264 (0.204 to 0.309)
	Overall NRI (95% CI)	0.494 (0.363 to 0.607)
Expert Reader CAC	Increase in LR chi-square	55.6
	Increase in AUC	0.033 (0.014 to 0.052)
	Event NRI (95% CI)	0.205 (0.120 to 0.294)
	Non-event NRI (95% CI)	0.298 (0.239 to 0.346)
	Overall NRI (95% CI)	0.503 (0.376 to 0.623)
Visually Estimated CAC	Increase in LR chi-square	32.0
	Increase in AUC	0.020 (0.007 to 0.033)
	Event NRI (95% CI)	0.174 (0.083 to 0.264)
	Non-event NRI (95% CI)	0.236 (0.177 to 0.298)
	Overall NRI (95% CI)	0.409 (0.278 to 0.537)

Supplemental Table 6: Net-reclassification analysis for the addition of coronary artery calcium (CAC) category. The reference model included all other components of the multivariable analysis outlined in Table 3. CI – confidence interval, NRI – net reclassification index.