Deep learning-based stenosis quantification from coronary CT Angiography

Youngtaek Hong, BS[a,b], **Frederic Commandeur, PhD**[b], **Sebastien Cadet, MS**[c], **Markus Goeller, MD**[b,d], **Mhairi K Doris, MD**[c,e], **Xi Chen, BSc**[c], **Jacek Kwiecinski, MD**[c,e], **Daniel S Berman, MD**[c], **Piotr J Slomka, PhD**[c], **Hyuk-Jae Chang, MD**[f], **Damini Dey, PhD**[b]

[a]Brain Korea 21 Project for Medical Science, Yonsei University, Seoul, Republic of Korea

[b]Biomedical Imaging Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[c]Departments of Imaging and Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[d]Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Faculty of Medicine, Department of Cardiology, Erlangen, Germany

[e]Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom

[f]Department of internal medicine, Division of Cardiology, Severance Cardiovascular Hosp, Yonsei University College of Medicine, Seoul, Korea, Republic of

Abstract

Background: Coronary computed tomography angiography (CTA) allows quantification of stenosis. However, such quantitative analysis is not part of clinical routine. We evaluated the feasibility of utilizing deep learning for quantifying coronary artery disease from CTA.

Methods: A total of 716 diseased segments in 156 patients (66 ± 10 years) who underwent CTA were analyzed. Minimal luminal area (MLA), percent diameter stenosis (DS), and percent contrast density difference (CDD) were measured using semi-automated software (Autoplaque) by an expert reader. Using the expert annotations, deep learning was performed with convolutional neural networks using 10-fold cross-validation to segment CTA lumen and calcified plaque. MLA, DS and CDD computed using deep-learning-based approach was compared to expert reader measurements.

Results: There was excellent correlation between the expert reader and deep learning for all quantitative measures (r=0.984 for MLA; r=0.957 for DS; and r=0.975 for CDD, p<0.001 for all). The expert reader and deep learning method was not significantly different for MLA (median 4.3 mm^2 for both, p=0.68) and CDD (11.6 vs 11.1%, p=0.30), and was significantly different for DS (26.0 vs 26.6%, p<0.05); however, the ranges of all the quantitative measures were within inter-observer variability between 2 expert readers.

Conclusions: Our deep learning-based method allows quantitative measurement of coronary artery disease segments accurately from CTA and may enhance clinical reporting.

Address for correspondence: Damini Dey, PhD, Associate Professor, Biomedical Imaging Research Institute, Cedars-Sinai Medical Center, 8700 Beverly Blvd., PACT, Suite 800, Los Angeles, CA 90048, Tel.: +1 310-423-1517, Damini.Dey@cshs.org.

Introduction

The improvement of CT image quality over the past decade allows direct evaluation of the entire coronary tree and assessment of both stenosis and coronary plaque from coronary CT Angiography (CCTA).[1–3] Clinical results from CCTA currently include only visual stenosis interpretation. Contrast density difference (CDD), a measure of luminal contrast kinetics, has been shown to be related to lesion-specific ischemia by invasive fractional flow reserve.[4] To date, quantitative measurements of stenosis or plaque from CCTA are not part of the clinical routine, due to the need for subjective and prohibitively time-consuming manual artery editing. A new class of machine learning algorithms, deep learning, including convolutional neural networks (CNN), has been shown to be very effective for automated object detection and image classification from a wide range of data.[5–10]

Purpose

Our objective was to evaluate the feasibility of segmentation of coronary lumen from CCTA using a CNN-based segmentation approach, and further, to evaluate quantitative stenosis, and other luminal image biomarkers such as minimum luminal area (MLA) and CDD, in comparison to expert readers.

METHODS

Patients and imaging protocol

In this retrospective study, we included 156 consecutive patients who underwent contrast-enhanced CCTA for clinical indications at Cedars-Sinai Medical Center. CCTA was performed on a dual-source 64-slice CT scanner (Definition Siemens Medical Solution, Forchheim, Germany) in accordance with societal guidelines. Beta-blockers (orally or intravenously) were administered prior to the scan. Prospective and retrospective gating protocols were utilized with a tube voltage of 100 kV in patients with a body mass index (BMI) < 30 kg/m^2 and 120 kV otherwise. Reconstruction parameters were: 512×512 matrix, 0.5×0.5 mm^2 pixel size, 0.6 mm slice thickness, and 0.3 mm slice increment. The study was approved by the Institutional Review Board, and all patients provided written informed consent.

Coronary CTA analysis

The reconstructed CCTA were transferred to a central database. Plaque analysis was performed on standard Windows workstations with the semi-automated software tool Autoplaque (version 2.0, Cedars-Sinai Medical Center, Los Angele, CA, USA)[11] by expert readers with level III or higher certification in cardiac CT. For each scan, the reader first placed a circular region-of-interest in the aortic root and scan-specific plaque thresholds were computed by the software as described previously[12, 13]. Quantitative measurements were made in each coronary segment using a standard 17-segment model[14] (SCCT guidelines) up to a luminal diameter limit   1.5 mm. Plaque quantification was performed with adaptive algorithms that are scan-specific, as previously described.[12, 13] Minimal lumen area (MLA) of the lesion were measured.[12, 13] Quantitative percent diameter stenosis (DS) was calculated by dividing the narrowest lumen diameter by the average of two normal non-

diseased reference cross-sections (proximal and distal to the lesion). CDD was defined as the maximum percent difference in contrast densities (luminal contrast density or attenuation per unit area), with respect to the proximal reference cross-section (with no disease). If the manual adjustments were needed, edits were made using the standardized correction options which allows for editing of vessel wall, lumen and adjustment of plaque thresholds. Interobserver variability was evaluated in a subset of 20 patients by two independent expert readers.

Data preparation for convolutional neural networks

The annotated segments were reconstructed to straightened views by using the extracted centerlines. The input of the convolutional networks were 2-D cross-sectional image, while the expert annotations were defined as the target references of the convolutional networks. We evaluated 2 sets of plaque thresholds for the input, defining 2 experiments: the scan-specific thresholds used in the software and a general threshold evaluated from a separate training cohort of 50 patients acquired at the same site with the same acquisition parameters (background <-10 HU, -10 to 200 HU for NCP, lumen 201 HU to 499 HU, Calcified Plaque (CP) 500 HU). In this study, the expert annotations had three labels as follows: background, contrast-enhanced lumen, and CP. We divided the target annotation into two types: 1) background and foreground (lumen + CP annotations); and 2) background and foreground (CP annotation only). Therefore, we prepared 2 sets of 4 channels input (CTA cross-sections preprocessed with the 4 HU threshold ranges), and 2 classes target (background and foreground). Two datasets were trained separately, one for each experiment.

Deep learning architecture

We designed a CNN architecture based on M-Net [15]. M-Net is an end-to-end CNN architecture initially proposed for segmenting brain structures from Magnetic Resonance Image, improving the standard U-shape CNN architecture U-Net[5, 6]. M-Net consists in a U-Net architecture, formed by a convolutional autoencoder, to which front and end pathways have been added to provide a deep supervision functionality. The convolutional architecture consists in two branches: the encoding composed by sequential units of two consecutive convolutional layers with a 3×3 kernel, and a 2×2 max-pooling to reduce the dimension; the decoding, symmetrically composed by the two convolutions and a 2×2 upsampling. Max-pooling and upsampling were also used in the front and end pathways, respectively, to ensure the transformation from the input resolution to the deepest layer, and inversely. A final convolutional layer with a 1×1 kernel, associated with a Softmax classifier, assigned to each pixel of the input one of the two classes: foreground or background. Parametric rectified linear activation and batch normalization were used to improve the training capacity and speed[16], and a dropout strategy was used with a 0.5 ratio during the training to limit overfitting[17]. We combined the binary cross-entropy and the Jaccard index, which measure the overlap between two segmentations, to optimize a joint loss function. The networks were trained with the Adam optimizer over 50 epochs.

Cross-validation

The entire experiments were performed using 10-fold cross-validation, for robust unbiased evaluation. During validation, all the lesion cases were randomly shuffled and split into 1/10 of the aggregate data. The model trained and validated with 9 subsets (578 lesions for training, 64 lesions for validation) and then tested against the remaining one subset (remaining 72 lesions). The testing was repeated 10 times and results concatenated until the entire dataset was segmented. MLA, DS, CDD were computed from both expert and deep learning for each lesion.

Statistical analysis

The primary end-point of this study was the performance of the deep learning method compared to the evaluation by the expert reader. Statistical analysis was performed using Excel add-in Analyse-it software (Analyse-it, Leeds, UK). Continuous variables were expressed as mean ± standard deviation or median and interquartile range (IQR). Deep learning and expert quantifications were systematically compared using the Spearman correlation coefficient, Bland-Altman plots and paired Wilcoxon rank-sum test. A p-value of < 0.05 was considered statistically significant. Deep learning performance was also evaluated with the Dice similarity coefficient (DSC), as a measure of overlap between expert and deep learning. This is computed by:

$$DSC = 2TP/(2TP+FP+FN),$$

Where TP (true positive) is the number of correctly positively classified voxels, FP (false positive) is the number of incorrectly positively classified, and FN (false negative) is the number of incorrectly negatively annotated classified voxels. The value of the DSC ranges from 0 to 1, where 0 means that there is no overlap and 1 means that there is complete overlap.

RESULTS

Quantitative CCTA measures

From 716 lesion segments, MLA, DS, and CDD were measured. After training, the typical time for computation was <32 seconds on a standard Windows workstation (mean 31.1 ± 21.0 seconds).

Figure 1 shows a representative case example of quantitative plaque analysis from our study. Figures 2 and 3 show the correlations and Bland-Altman plots for the quantitative measures DS and CDD. Overall, both general threshold and scan-specific threshold quantifications showed excellent correlation and agreement for MLA, DS and CDD with expert quantification (Table 2). The scan-specific threshold quantification demonstrated Spearman rank coefficients from 0.96 to 0.99, and the general threshold quantification showed rank coefficients from 0.86 to 0.91.

Table 1 shows the median and IQR of the imaging biomarkers with the p-values for pairwise comparison between deep learning and expert quantification. The correlation coefficients

and Bland-Altman 95% confidence intervals for the comparison of quantitative plaque measures using general or scan-specific threshold are shown in Table 1. While the p-values show significant differences for all the biomarkers for the general thresholds, MLA and CDD were not significantly different from expert quantification for scan-specific threshold only (P=0.6786 and P=0.2996, respectively). From the Bland-Altman plots in Figures 2 and 3, for all the quantitative measures, the 95% limits of agreement for the general thresholds were wider by 200–300%, indicating worse agreement with the expert reader (Table 1). Even for scan-specific thresholds, few outliers were observed, associated to ostial lesions particularly at the ostium of the Left Main artery, for which the performance of deep learning was lower/less accurate. Table 2 shows the absolute difference between 2 expert readers for the quantitative parameters. For all the quantitative parameters obtained with scan-specific thresholds, 95% limits of agreement for the deep learning method were within range of the differences between 2 expert readers. A resulting DSC of 0.95±0.01 for the lumen was obtained, showing high accuracy in the segmentation.

New/breakthrough work to be presented

In this study, we evaluated the feasibility of deep learning for quantitative coronary artery segmentation and showed that clinically relevant parameters such as MLA, DS and CDD can be accurately measured from CCTA. Our deep learning model successfully captures intrinsic features of contrast-enhanced lumen and CP from CCTA. To our knowledge, this has not been demonstrated before. We also evaluated two sets of plaque thresholds for measurement of MLA, DS and CDD. These two methods showed a strong correlation with expert quantification, with scan-specific thresholds yielding more accurate results. While this requires minor manual interactions with semi-automated software (placing of a standard region-of-interest at the aortic root), these are not time-consuming tasks and with further development of this new approach, could potentially become fully automated.

CONCLUSIONS

Our deep learning-based method allows accurate quantitative measurement of coronary artery disease, and may enhance clinical reporting.

Acknowledgements/Funding:

Abbreviations
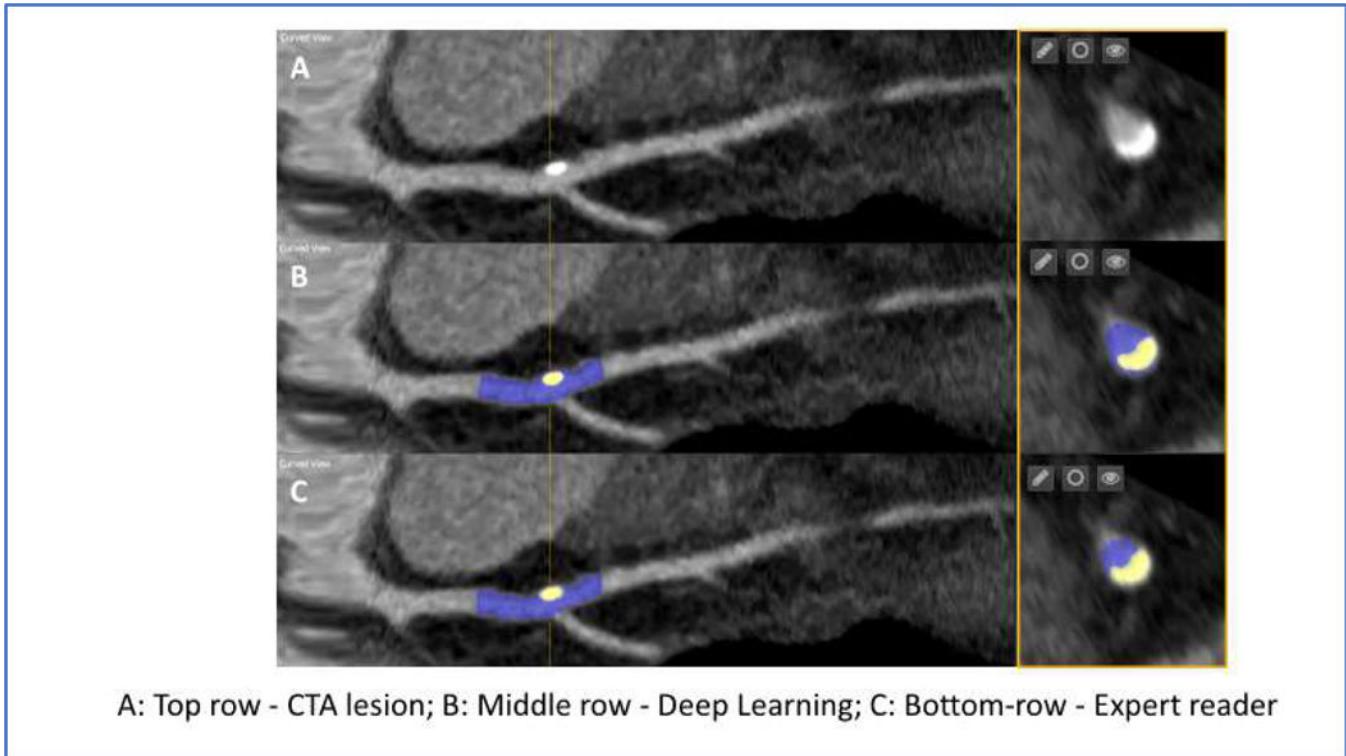
CCTA	Coronary computed tomographic angiography
CAD	Coronary artery disease
CDD	Contrast density difference
CNN	Convolutional neural networks
CP	Calcified plaque

MLA	Minimal lumen area
DS	Diameter stenosis
IQR	Interquartile range
DSC	Dice similarity coefficient

References
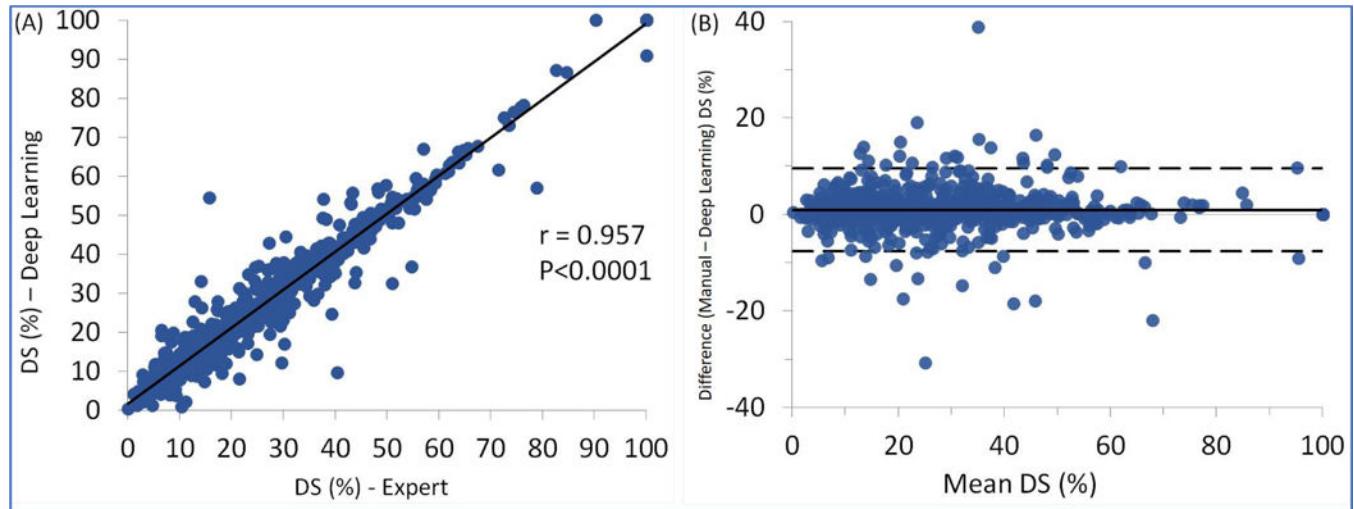
1. Gaur S, Ovrehus KA, Dey D, Leipsic J, Botker HE, Jensen JM, Narula J, Ahmadi A, Achenbach S, Ko BS, Christiansen EH, Kaltoft AK, Berman DS, Bezerra H, Lassen JF and Norgaard BL. Coronary plaque quantification and fractional flow reserve by coronary computed tomography angiography identify ischaemia-causing lesions. Eur Heart J 2016;37:1220–7. [PubMed: 26763790]

2. Hell M, Motwani M, Otaki Y, Cadet S, Gransar H, Miranda-Peats R, Valk J, SLomka PJ, Cheng V, Rozanski R, Tamarappoo BK, Hayes S, Achenbach S, Berman DS, Dey D. Quantitative global plaque characteristics from coronary CT Angiography for the prediction of future cardiac death during 5 years of follow-up. European Heart Journal Cardiovascular Imaging 2017;18:1331–1339. [PubMed: 28950315]

3. Diaz-Zamudio MFT, Slomka PJ, Arsanjani R, Gransar H, Germano G, Berman DS, Kaufmann P, Dey D. Quantitative plaque features from coronary computed tomography angiography to identify regional ischemia by myocardial perfusion imaging Eur Heart J Cardiovasc Imaging 2017;18:499–507. [PubMed: 28025263]

4. Dey D, Gaur S, Ovrehus KA, Slomka PJ, Betancur J, Goeller M, Hell MM, Gransar H, Berman DS, Achenbach S, Botker HE, Jensen JM, Lassen JF and Norgaard BL. Integrated prediction of lesion-specific ischaemia from quantitative coronary CT angiography using machine learning: a multicentre study. European radiology 2018.

5. Ronneberger O, Fischer P and Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. 2015:234–241.

6. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T and Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention. 2016:424–432.

7. Commandeur F, Goeller M, Betancur J, Cadet S, Doris M, Chen X, Berman DS, Slomka PJ, Tamarappoo BK and Dey D. Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue from Non-Contrast CT. IEEE transactions on medical imaging 2018;PP:1–1. [PubMed: 28945591]

8. Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, Leiner T and Išgum I. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. Medical image analysis 2018;44:72–85. [PubMed: 29197253]

9. Beecy A, Chang Q, Baskaran L, Elmore K, Kolli K, Hao Wang, Anchouche K, Al'Aref S, Pena J, Knight-Greenfield A, Patel P, Sun P, Zhang T, Kamel H and Gupta A. A Novel Deep Learning Approach for Automated Diagnosis of Cerebrovascular Accidents; 2017.

10. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, Fish MB, Ruddy TD, Kaufmann P, Sinusas AJ, Miller EJ, Bateman TM, Dorbala S, Di Carli M, Germano G, Otaki Y, Tamarappoo BK, Dey D, Berman DS and Slomka PJ. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT. A Multicenter Study 2018.

11. Dey D, Cheng VY, Slomka PJ, Nakazato R, Ramesh A, Gurudevan S, Germano G and Berman DS. Automated 3-dimensional quantification of noncalcified and calcified coronary plaque from coronary CT angiography. Journal of cardiovascular computed tomography 2009;3:372–82. [PubMed: 20083056]

12. Dey D, Cheng VY, Slomka PJ, Nakazato R, Ramesh A, Gurudevan S, Germano, Berman DS. Automated Three-dimensional Quantification of Non-calcified and Calcified Coronary Plaque

from Coronary CT Angiography. Journal of Cardiovascular Computed Tomography 2009:372–382. [PubMed: 20083056]

13. Dey D, Schepis T, Marwan M, Slomka PJ, Berman DS, Achenbach S. Automated Three-dimensional Quantification of Non-calcified Coronary Plaque from Coronary CT Angiography: comparison with Intravascular Ultrasound Radiology 2010;257:516–522. [PubMed: 20829536]

14. Leipsic J, Abbara S, Achenbach S, Cury R, Earls JP, Mancini GBJ, Nieman K, Pontone G and Raff GL. SCCT guidelines for the interpretation and reporting of coronary CT angiography: A report of the Society of Cardiovascular Computed Tomography Guidelines Committee. Journal of cardiovascular computed tomography 8:342–358. [PubMed: 25301040]

15. Mehta R and Sivaswamy J. M-net: A Convolutional Neural Network for deep brain structure segmentation. Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. 2017:437–440.

16. Ioffe S and Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning. 2015:448–456.

17. Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014;15:1929–1958.

A: Top row - CTA lesion; B: Middle row - Deep Learning; C: Bottom-row - Expert reader
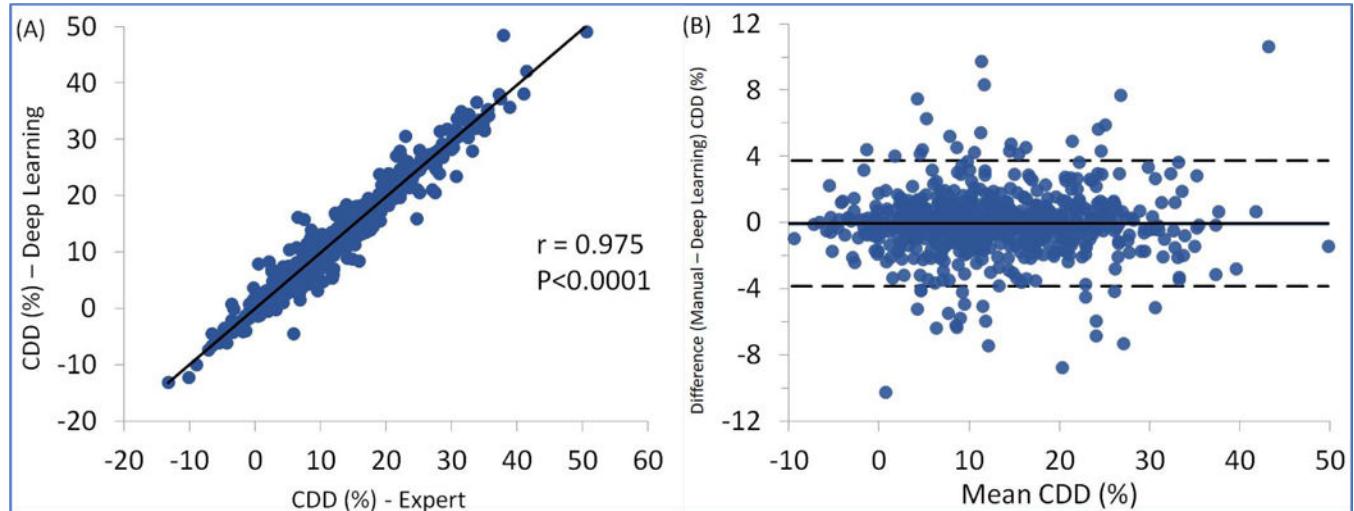
Figure 1.
Representative case example of lumen and calcified plaque segmentation in a CCTA lesion extending from left main to left anterior descending artery (LM-LAD). (A) Top row: CCTA lesion - Curved planar reformatted view and cross-sectional view. (B) Middle row: Deep learning-based approach (C) Bottom row: Expert reader annotation. Blue indicates contrast-enhanced lumen and yellow indicates calcified plaque.

Figure 2.
(A) Spearman correlation and (B) Bland-Altman plots for percent diameter stenosis, between deep learning (using scan-specific thresholds) and expert quantification. Excellent correlation was observed r = 0.957. The 95% of limits agreements were −7.7 to 9.5%.

Figure 3.
(A) Spearman correlation and (B) Bland-Altman plots for percent contrast density difference, between deep learning (using scan-specific thresholds) and expert quantification. Excellent correlation was observed r = 0.975, and 95% of limits agreements were −3.9 to 3.8 %.

Table 1

Comparison of deep learning vs expert (median and interquartile ranges are reported for each measure)

Quantitative parameter	Plaque thresholds utilized	
	General	Scan-specific
Minimal Lumen Area (mm2)		
Deep Learning Evaluation	*4.2 (2.5 – 6.4)*	4.4 (2.6 – 6.6)
Expert (manual) Evaluation	*4.4 (2.6 – 6.6)*	4.4 (2.6 – 6.6)
P – value	*0.0043*	0.6786[+]
Correlation (Spearman rank)	*0.915*	0.984
Bland Altman (95% limits; range)	*−2.7, 2.3; 5.0*	−1.3, 1.3; 2.6
Diameter Stenosis (%)		
Deep Learning Evaluation	*28.2 (17.3 – 39.6)*	26.6 (16.3 – 38.5)
Expert (manual) Evaluation	*26.0 (15.1 – 37.9)*	26.0 (15.1 – 37.9)
P – value	*<0.0001*	<0.0001
Correlation (Spearman rank)	*r = 0.873*	r = 0.957
Bland Altman (95% limits; range)	*−18.3, 23.0; 41.3*	−7.7, 9.5; 17.2
Contrast density difference (%)		
Deep Learning Evaluation	*13.2 (7.6 – 21.4)*	11.2 (5.6 – 18.4)
Expert (manual) Evaluation	*11.6 (6.0 – 18.3)*	11.6 (6.0 – 18.3)
P – value	*<0.0001*	0.2996[+]
Correlation (Spearman rank)	*r = 0.873*	r = 0.975
Bland Altman (95% limits; range)	*−8.0, 12.0; 20.0*	−3.9, 3.8; 7.7

Generic indicates that general threshold applied CNN model, Scan-specific indicates that scan-specific threshold applied CNN model.

Median (1st quartile – 3rd quartile), P-value is Wilcoxon rank sum test and $p < 0.05$ is considered as significant.

[+]For MLA and CDD, pairwise differences were significantly not different for the scan-specific method, indicating equivalence.

Table 2

Inter-observer variability (absolute difference) in quantitative measures

Quantitative parameters	
Minimal Lumen Area (mm^2)	
Mean	1.5 ± 2.2
95% CI; range	0 – 6.3; 6.3
Diameter Stenosis (%)	
Mean	11.6 ± 11.8
95% CI; range	1.0 – 36.0; 35.0
Contrast density difference (%)	
Mean	2.6 ± 2.6
95% CI	0 – 7.9; 7.9

For all quantitative parameters, 95% limits of agreement for the deep learning method were within range of the differences between 2 expert readers.