

Calcium Scoring at Coronary CT Angiography Using Deep Learning

Dan Mu, PhD* • Junjie Bai, PhD* • Wenping Chen, MS • Hongming Yu, MS • Jing Liang, MS • Kejie Yin, MS • Hui Li, MS • Zhao Qing, PhD • Kelei He, PhD • Hao-Yu Yang, MS • Jinyao Zhang, MD • Youbing Yin, PhD • Hunter W. McLellan, BS • U. Joseph Schoepf, MD • Bing Zhang, MD, PhD

From the Department of Radiology, Affiliated Nanjing Drum Tower Hospital of Nanjing University Medical School, Nanjing, China (D.M., W.C., H.Y., J.L., K.Y., H.L., Z.Q., B.Z.); Keya Medical, Shenzhen, China (J.B., H.Y.Y., J.Z., Y.Y.); Medical School of Nanjing University, Nanjing, China (K.H.); National Institutes of Healthcare Data Science at Nanjing University, Nanjing, China (K.H.); University of South Carolina School of Medicine–Columbia, Columbia, SC (H.W.M.); Division of Cardiovascular Imaging, Medical University of South Carolina, Charleston, SC (U.J.S.); Institute of Brain Science, Nanjing University, Nanjing 210008, China (B.Z.). Received June 13, 2021; revision requested July 28; revision received September 10; accepted September 28. Address correspondence to B.Z. (e-mail: zhangbing_nanjing@nju.edu.cn).

This work was supported by the National Natural Science Foundation of China (grant no. 81720108022 for B.Z.; grant no. 81601539 for D.M.); the Fundamental Research Funds for the Central Universities, Nanjing University (grant no. 2020-021414380462); the key project of Jiangsu Commission of Health (grant no. K2019025); the social development project of science and technology project in Jiangsu Province (grant no. BE2017707); the key medical talents of the Jiangsu province, the 13th Five-Year health promotion project of the Jiangsu province (grant no. ZDRCA2016064); Jiangsu Provincial Key Medical Discipline (Laboratory) (grant no. ZDXKA2016020); the Six Talent Peaks Project in Jiangsu Province (grant no. WSN-138); the Nanjing Medical Science and Technique Development Foundation (grant nos. ZKX19018 and QRX17057); China Postdoctoral Science Foundation (grant no. 2019M661804); Jiangsu Province Postdoctoral Science Foundation (grant no. 2019K060); and Shenzhen Science and Technology Innovation Program (grant No. KQTD2016112809330877).

* D.M. and J.B. contributed equally to this work.

Conflicts of interest are listed at the end of this article.

See also the editorial by Goldfarb and Cao et al in this issue.

Radiology 2021; 000:1–8 • <https://doi.org/10.1148/radiol.2021211483> • Content codes: **CA** **CT**

Background: Separate noncontrast CT to quantify the coronary artery calcium (CAC) score often precedes coronary CT angiography (CTA). Quantifying CAC scores directly at CTA would eliminate the additional radiation produced at CT but remains challenging.

Purpose: To quantify CAC scores automatically from a single CTA scan.

Materials and Methods: In this retrospective study, a deep learning method to quantify CAC scores automatically from a single CTA scan was developed on training and validation sets of 292 patients and 73 patients collected from March 2019 to July 2020. Virtual noncontrast scans obtained with a spectral CT scanner were used to develop the algorithm to alleviate tedious manual annotation of calcium regions. The proposed method was validated on an independent test set of 240 CTA scans collected from three different CT scanners from August 2020 to November 2020 using the Pearson correlation coefficient, the coefficient of determination, or r^2 , and the Bland-Altman plot against the semiautomatic Agatston score at noncontrast CT. The cardiovascular risk categorization performance was evaluated using weighted κ based on the Agatston score (CAC score risk categories: 0–10, 11–100, 101–400, and >400).

Results: Two hundred forty patients (mean age, 60 years \pm 11 [standard deviation]; 146 men) were evaluated. The positive correlation between the automatic deep learning CTA and semiautomatic noncontrast CT CAC score was excellent (Pearson correlation = 0.96; r^2 = 0.92). The risk categorization agreement based on deep learning CTA and noncontrast CT CAC scores was excellent (weighted κ = 0.94 [95% CI: 0.91, 0.97]), with 223 of 240 scans (93%) categorized correctly. All patients who were miscategorized were in the direct neighboring risk groups. The proposed method's differences from the noncontrast CT CAC score were not statistically significant with regard to scanner (P = .15), sex (P = .051), and section thickness (P = .67).

Conclusion: A deep learning automatic calcium scoring method accurately quantified coronary artery calcium from CT angiography images and categorized risk.

© RSNA, 2021

Coronary artery calcium (CAC) scoring is an effective risk stratification tool for coronary artery disease (1). Multiple types of calcium scoring exist, among which, to our knowledge, the Agatston calcium score is the most widely used (2). Traditionally, the Agatston score is based on noncontrast CT scans. Voxels within the course of a coronary artery that display an attenuation exceeding a predefined threshold of more than 130 HU are defined as calcium. The score is then computed by means of a predefined equation (2). In the past, a semiautomatic approach was adopted, in which human interaction was needed to differentiate coronary calcium from calcium deposits in other locations, such as the aorta and bones, at noncontrast CT. Recently, attempts have been made to automatically calculate Agatston

scores at noncontrast CT (3–10), including deep learning–based methods (3,5,8,10,11).

CT angiography (CTA), using an injected contrast material to enhance the coronary arteries, is a common non-invasive modality to evaluate the anatomic and functional information of the coronary artery tree (12,13). Ordinarily, CTA is preceded by a noncontrast CT CAC scoring study. It is desirable to evaluate CAC scores directly at CTA, as it eliminates the additional radiation associated with the need for extra noncontrast CT imaging. However, the finding of automatic CAC scoring at CTA, which is equivalent to the widely used Agatston scores for noncontrast CT, has not previously been well established in the literature. Because of the hyperattenuating appearance of a contrast material–filled artery, it is challenging to extract the CAC region at CTA.

Abbreviations

CAC = coronary artery calcium, CTA = CT angiography, ICC = intraclass correlation coefficient, VNC = virtual noncontrast

Summary

An automatic deep learning approach accurately quantified calcium scores at coronary CT angiography and was validated against noncontrast CT coronary artery calcium scores using multiple imaging protocols.

Key Results

- In this retrospective study of 240 patients undergoing coronary CT angiography, the correlation between the deep learning calcium scoring and noncontrast CT coronary artery calcium (CAC) score was excellent ($r^2 = 0.92$, $P < .001$).
- The risk categorization agreement between deep learning and noncontrast CT CAC was excellent (weighted $\kappa = 0.94$), with 223 of 240 patients (93%) categorized correctly.
- The proposed method's differences from the noncontrast CT CAC score in the test set were not significantly different with regard to scanner ($P = .15$), sex ($P = .051$), and section thickness ($P = .67$).

Typically, a threshold is used to define the CAC region; for example, a scan-specific threshold is heuristically determined using either an aorta Hounsfield unit (14,15) or a coronary artery Hounsfield unit such as fitting a trendline (16) or analyzing histogram (17) along a coronary artery. These heuristic thresholding schemes are usually validated in relatively small patient populations using a single imaging protocol (14,15,17), which

may not have been robust enough to handle larger variations. Previous attempts to derive CAC scores from CTA using segmentation-based deep learning on small data sets have been conducted; however, they relied on manual voxel-level calcification annotation (18). Manual annotation is not only tedious, but it is also prone to error. Additionally, such segmentation methods cannot directly provide the widely used Agatston score and the corresponding risk categories. This is because of the attenuation-modulated weighted summation no longer being applicable as the attenuation values may vary with imaging protocols.

Spectral CT exploits the energy dependence of x-ray photon attenuation, providing more detailed tissue analysis based on the material decomposition under multiple energy levels (19). By using the material decomposition virtual iodine subtraction algorithm, spectral CT scanners are able to generate virtual noncontrast (VNC) CT scans. Also, without requiring extra radiation, postimaging reconstructions may replace true noncontrast CT (20,21).

In this study, a deep learning method was proposed to automatically quantify CAC scores at CTA. In the training stage, VNC scans, alongside the CTA scans obtained with spectral CT, were used to efficiently obtain the reference calcium region masks in the training set using an automatic initial mask followed by expert review. The model was comprehensively evaluated on an independent testing set with multiple imaging protocols. VNC scans were used to develop the algorithm; however, the final algorithm was tested on contrast-enhanced coronary CTA scans without spectral postprocessing.

Materials and Methods

The ethics committee of Nanjing Drum Tower Hospital, a major teaching hospital affiliated with Nanjing University Medical School in Nanjing, China, approved the study, and the need to obtain informed written consent for participation was waived.

Patient Data Sets

This retrospective study was designed to evaluate a deep learning CAC score quantification from a single CTA scan against noncontrast CT CAC scoring. The study included a training and validation set of patients who had undergone spectral CT and an independent test set of patients who had undergone spectral CT or conventional single-energy CT (Fig 1). Exclusion criteria included scans with existing stents and scans with severe artifacts (ie, motions, missing sections, etc). The reference semiautomatic noncontrast CT CAC score was computed using the Agatston method (2). Patients were stratified

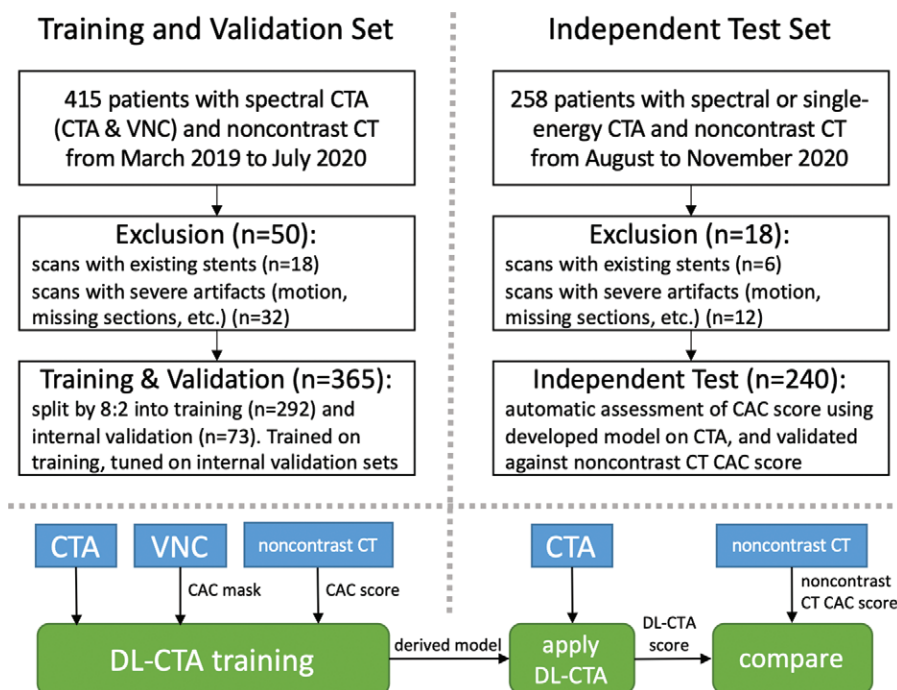


Figure 1: Flowchart of training and validation set and independent test set. For method development, CT angiography (CTA) and virtual noncontrast (VNC) CT scans from spectral CT scanner and noncontrast CT scans in 415 patients were obtained from March 2019 to July 2020. After applying exclusion criteria, training and validation set ($n = 365$) was used for training and parameter tuning. Independent test set ($n = 240$) with CTA scans from three different spectral or single-energy scanners and noncontrast CT scans were obtained from August 2020 to November 2020. This was used to validate derived method after applying same exclusion criteria. CAC = coronary artery calcium, DL = deep learning.

into four different risk categories (ie, 0–10, 11–100, 101–400, and >400), according to the Agatston score (6,15).

The training and validation sets ($n = 365$) and test set ($n = 240$), which were randomly selected from those who underwent both CTA and noncontrast CT examinations, were obtained temporally independently (ie, from March 2019 to July 2020 and from August 2020 to November 2020, respectively) from Nanjing Drum Tower Hospital. Figure 1 shows how data sets were obtained while applying exclusion criteria, training, and validation.

Imaging Protocol

Coronary CT scans were obtained following standard imaging protocols. The CTA and noncontrast CT scans from each patient were obtained at the same imaging session. All images in the training and validation set were obtained with a spectral CT scanner (IQon Spectral CT, Philips). The tube voltage range was 120–140 kV, and the tube current range was 148–752 mA. Images in the temporally independent test set were obtained using the IQon Spectral CT scanner (Philips), the iCT 256 scanner (Philips), or the Revolution CT scanner (GE Healthcare). The tube voltage range was 100–140 kV, and the tube current range was 163–992 mA. For CTA imaging, 50 mL of iopromide (370 mg/mL; Ultravist, Bayer) was injected at a flow rate of 4–5 mL/sec.

Semiautomatic Calcium Scoring at Noncontrast CT and VNC CT

The reference Agatston scores at noncontrast CT were semiautomatically obtained by an experienced radiologist (reader 1, J.L., with more than 8 years of experience) using the IntelliSpace Portal (version 10.0, Philips). To study the intra- and interobserver variability, another radiologist (reader 2, W.C., with more than 8 years of experience) was asked to evaluate all 240 noncontrast CT scans in the test set with an interval of 2 weeks. The mean of the two readers' first scoring was used as the reference noncontrast CT CAC score. Addition-

ally, calcium scores from the VNC images in the test set ($n = 100$) CT were semiautomatically obtained for comparison.

Automatic Calcium Scoring at CTA Using Deep Learning

A deep learning method was proposed to take a CTA scan, and from it, to generate a calcium segmentation and corresponding CAC score (Fig 2). It consists of a calcium detection module and a CAC score regression module, both using three-dimensional convolutional neural networks. The calcium detection module generates a calcium region mask using a multiresolution encoder and decoder structure with skip connections, similar to U-Net (22). The CAC score regression module uses features from multiple scales to regress the corresponding CAC score.

During training, calcium region masks corresponding to the CTA scans were required. To avoid the labor-intensive and error-prone task of manually annotating calcium regions, VNC scans obtained at spectral CT alongside CTA were used to obtain the initial calcium region masks (Fig 3). First, candidate CAC regions on VNC scans were demarcated as regions higher than 130 HU. Second, coronary artery segmentation was automatically performed at CTA using a deep learning algorithm (23–25). As the VNC scan had perfect alignment with CTA, only candidate CAC regions within a 3-mm proximity to the coronary artery were kept as a CAC region mask. One radiologist (J.L.) reviewed the calcium region mask for quality assurance as well as exclusion of aortic calcium near the ostia.

The reference CAC region mask, alongside the reference Agatston score obtained at noncontrast CT, was used in conjunction to train the proposed deep learning model in the training and validation set. Once trained, the model could be applied to CTA scans without VNC or noncontrast CT in the test set to directly quantify CAC.

The code to reproduce the results in this study is available at https://github.com/deronmontal/dl_cta_calcium.

Statistical Analysis

The trained deep learning model's performance was evaluated using the independent test set, which had no time overlap with

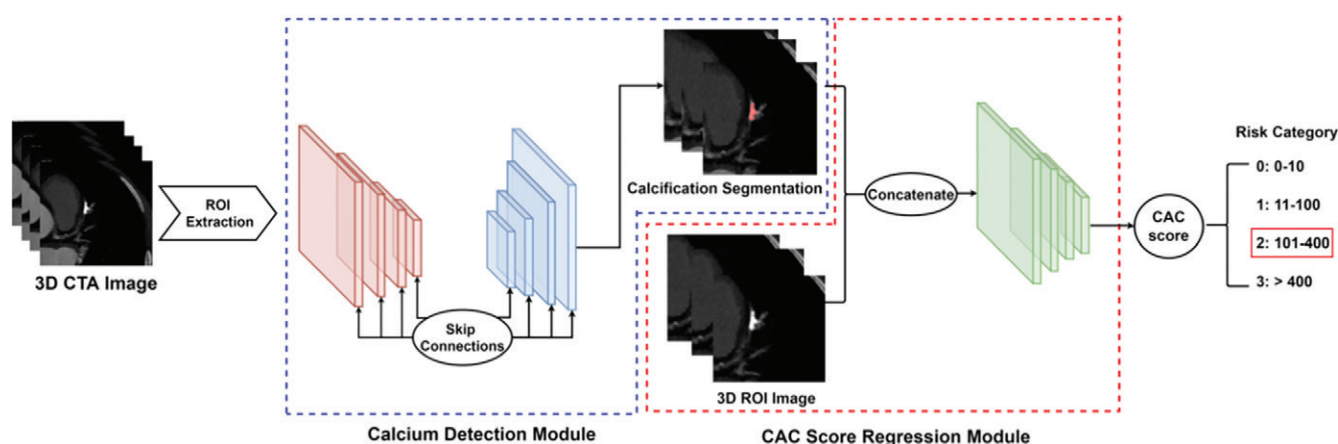


Figure 2: Diagram of proposed deep learning method for coronary artery calcium (CAC) quantification from CT angiography (CTA) scans. Calcium detection module detects calcification from CTA scans, and the CAC score regression module estimates CAC scores based on CTA scan and calcium detection results. Risk categorization is then performed based on estimated CAC score. ROI = region of interest, 3D = three-dimensional.

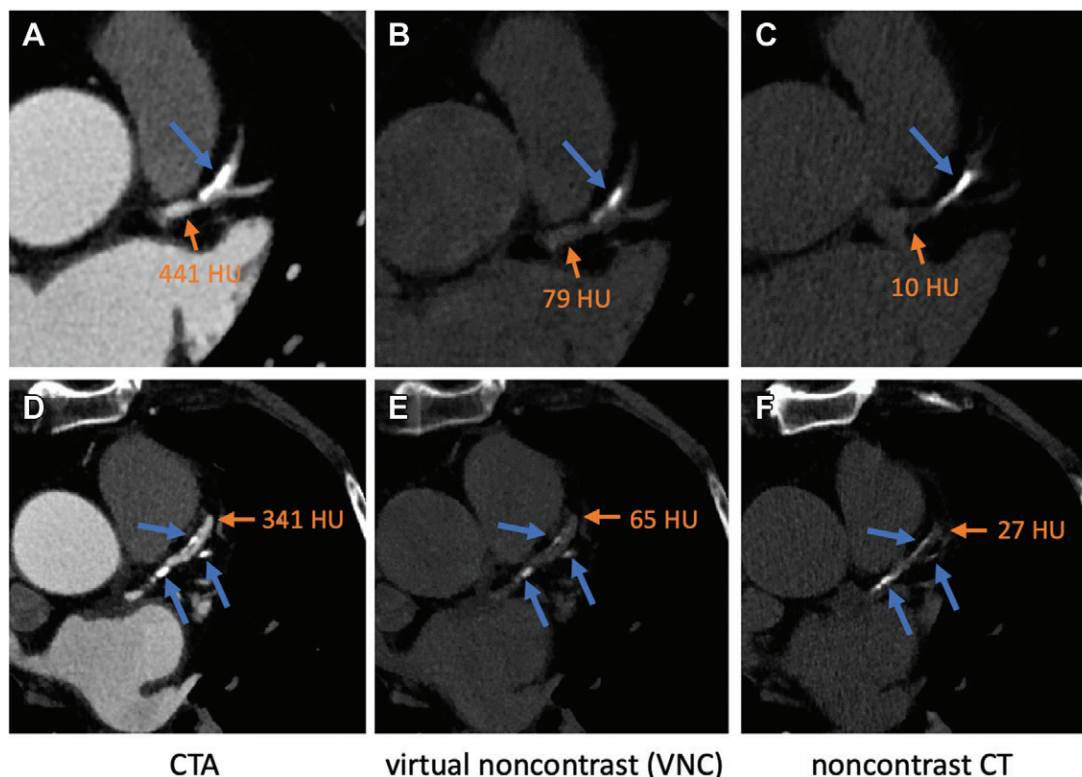


Figure 3: (A–F) Calcifications (blue arrows) are shown on CT angiography (CTA), VNC CT, and noncontrast CT scans in two patients in training set. Appearance of calcifications on VNC CT scans (B, E) and noncontrast CT scans (C, F) is similar and can be easily obtained using predefined threshold of 130 HU, as coronary artery attenuation (orange arrow in B, C, E, and F) is clearly smaller than 130 HU. However, obtaining coronary artery calcium (CAC) region and estimating corresponding CAC score by means of CTA (A, D) can be challenging because of variable attenuation of contrast-enhanced artery.

the training and validation sets. Statistical analysis was performed with R software (version 4.0.2, R Foundation for Statistical Computing), and statistical significance was defined as two-sided $P < .05$. Analysis of variance tests and Fisher exact tests were used to compare the differences among the different groups within the continuous and dichotomous variables, respectively. Continuous variables were expressed as means with standard deviations, whereas categorical variables were expressed as frequencies with percentages.

The intra- and interobserver repeatability of CAC scoring at noncontrast CT and VNC CT was analyzed using the intraclass correlation coefficient (ICC) and root mean square error. The correlation between the proposed method and the reference Agatston score was assessed using the Pearson correlation coefficient, the coefficient of determination, or r^2 , and ICC. The agreement was evaluated using the Bland-Altman plot. Because errors tended to be higher with greater CAC scores, the regression method for nonuniform differences was used to estimate the 95% limits of agreement as in van Velzen et al (10) and Sevrukov et al studies (26). The agreement of risk categorization was evaluated using weighted κ .

Results

Characteristics of Patient Data Sets

Table 1 lists the baseline patient information for the training and validation set and test set, respectively. A total of 365

patients (mean age, 57 years \pm 11 [standard deviation]; 207 men) were included in the final training and validation set. Both VNC and CTA scans, which had identical spatial information with perfect alignment, were reconstructed with a spacing of 0.32–0.66 mm, a section thickness of 0.8–1.0 mm, and a section interval of 0.4–0.45 mm. The test set contained a total of 240 patients (mean age, 60 years \pm 11; 146 men), with 100 scanned with the IQon Spectral CT scanner (Philips), 100 scanned with the iCT 256 scanner (Philips), and 40 scanned with the Revolution CT scanner (GE Healthcare). The CTA scans were reconstructed with a spacing of 0.33–0.71 mm, a section thickness of 0.63–0.9 mm, and a section interval of 0.45–0.63 mm. The noncontrast CT scans were reconstructed with a spacing of 0.31–0.98 mm, a section thickness of 2.5 mm, and a section interval of 2.5 mm. The dose-length product for noncontrast CT and CTA was 49.3 mGy \cdot cm \pm 8.6 and 462.1 mGy \cdot cm \pm 85.7, respectively. The estimated radiation dose for noncontrast CT and CTA was 0.7 mSv \pm 0.1 and 6.5 mSv \pm 1.2, respectively.

Intra- and Interobserver Variability of CAC Scoring at Noncontrast CT and VNC CT

Table 2 shows the intra- and interobserver variability of semi-automatic CAC scoring at noncontrast CT and VNC CT, respectively. The intraobserver variability was measured between reader 2's two readings. The interobserver variability was measured between reader 1's and reader 2's first reading for simplic-

Table 1: Characteristics of Patient Data Sets

Parameter	All (<i>n</i> = 605)	Training and Validation Set (<i>n</i> = 365)	Independent Test Set (<i>n</i> = 240)
Age (y)	58 ± 11	57 ± 11	60 ± 11
No. of men*	353 (58)	207 (57)	146 (61)
Scanner used	IQon Spectral CT (Philips), iCT 256 (Philips), and Revolution (GE)	IQon Spectral CT (Philips)	IQon Spectral CT (Philips), iCT 256 (Philips), and Revolution (GE)
Tube voltage (kV)	117 ± 8	120 ± 2	112 ± 10
Tube current (mA)	349 ± 137	296 ± 91	431 ± 155
Section thickness (mm)	0.88 ± 0.07	0.90 ± 0.01	0.85 ± 0.10

Note.—Except where indicated, data are means ± standard deviations.

* Numbers in parentheses are percentages.

Table 2: Intra- and Interobserver Variability of Semiautomatic CAC Scoring on Noncontrast CT and VNC Scans in Test Set

Observer Type	No. of Patients	Readings Compared	Continuous Variable*	ICC [†]	RMSE
NCCT intraobserver	240	R2 (first reading) vs R2 (second reading)	0.1 ± 4	1 (1, 1)	4
NCCT interobserver	240	R1 vs R2 (first reading)	40 ± 185	0.96 (0.95, 0.97)	188
VNC intraobserver	100	R2 (first reading) vs R2 (second reading)	0.2 ± 1	1 (1,1)	1
VNC interobserver	100	R1 vs R2 (first reading)	1 ± 7	1 (1,1)	7

Note.—All patients in the test set (*n* = 240) had noncontrast CT scans, among which only scans obtained with the IQon Spectral CT scanner (Philips) (*n* = 100) had virtual noncontrast (VNC) scans. CAC = coronary artery calcium, ICC = intraclass correlation coefficient, NCCT = noncontrast CT, RMSE = root mean square error, R1 = reader 1 (J.L.), R2 = reader 2 (W.C.).

* Numbers are means ± standard deviations.

[†] Numbers in parentheses are 95% CIs.

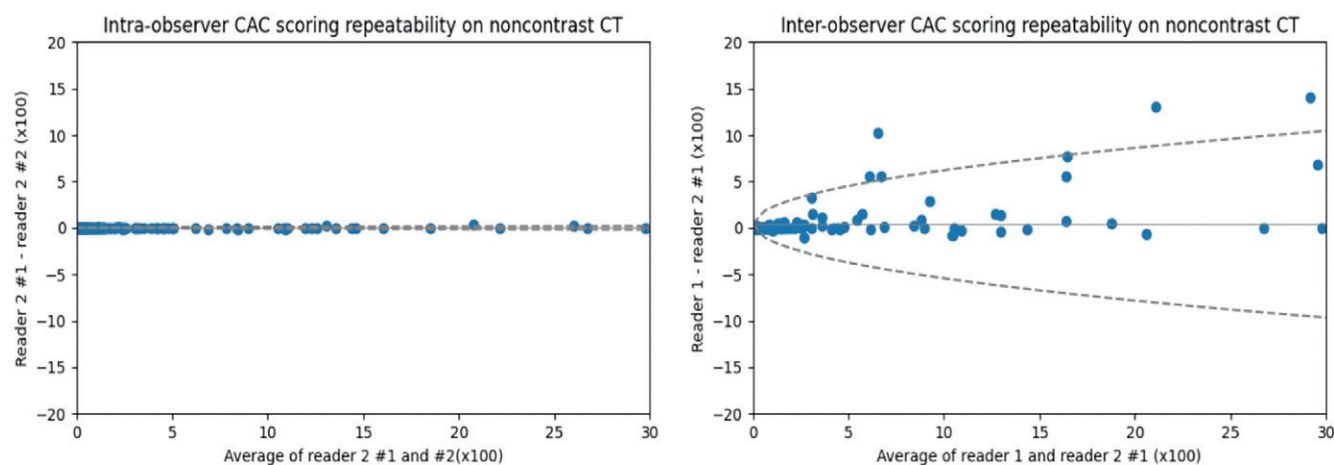


Figure 4: Bland-Altman plots of intra- and interobserver repeatability of semiautomatic Agatston scoring at noncontrast CT. The equation, difference = $\pm 1.96 \cdot (\pi/2)^{0.5} \cdot (b + a \cdot \text{mean}^{0.5})$, represented 95% limits of agreement. For intraobserver difference, *a* and *b* were 0.15 and 0.05, respectively. For interobserver difference, *a* and *b* were 7.5 and 16.3, respectively. CAC = coronary artery calcium.

ity, as reader 2's two readings were quite close. The ICCs were 1 (95% CI: 1, 1) and 0.96 (95% CI: 0.95, 0.97) for the intra- and interobserver variability at noncontrast CT semiautomatic reading, respectively. Both showed great consistency with slightly better consistency for intraobserver readings. Agreement of intra- and interobserver readings at noncontrast CT is shown in the Bland-Altman plot in Figure 4. The 95% limits of agreement were represented by the following equation: difference = $\pm 1.96 \cdot (\pi/2)^{0.5} \cdot (b + a \cdot \text{mean}^{0.5})$. For intraobserver

distance, *a* and *b* were 0.15 and 0.05, respectively. For interobserver difference, *a* and *b* were 7.5 and 16.3, respectively.

In addition, the ICC was 1 (95% CI: 1, 1) for both the intra- and interobserver variability of VNC semiautomatic reading with a root mean square error of 7 or less, showing nearly perfect consistency. The positive correlation between semiautomatic CAC scoring obtained at noncontrast CT and VNC CT was excellent (Pearson correlation coefficient = 0.95 [95% CI: 0.93, 0.97], *P* < .001; *r*² = 0.90, *P* < .001).

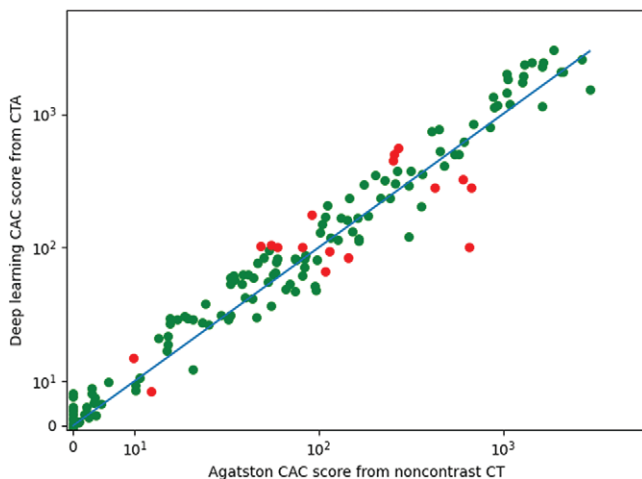


Figure 5: Scatterplot shows deep learning CT angiography (CTA) coronary artery calcium (CAC) score and Agatston noncontrast CT CAC scores in independent test set. Green dots represent risk categorization agreement, and red dots represent risk categorization disagreement.

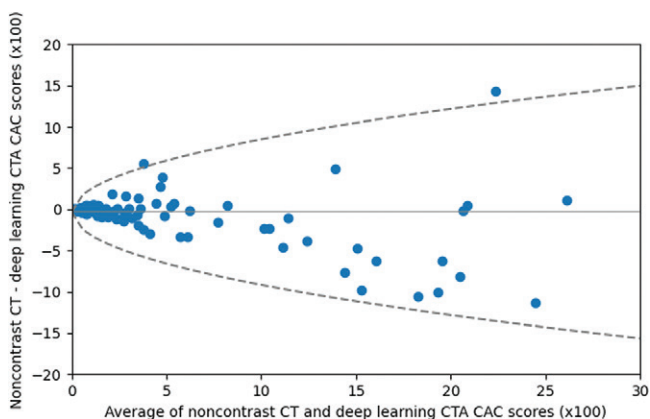


Figure 6: Bland-Altman plot shows Agatston coronary artery calcium (CAC) score at noncontrast CT and deep learning CAC score at CT angiography (CTA) in test set. The equation, difference = $\pm 1.96 \cdot (\pi/2)^{0.5} \cdot (b + a \cdot \text{mean}^{0.5})$, represented 95% limits of agreement, where a and b were 11.4 and -13.3 , respectively.

Positive Correlation

The positive correlation between the deep learning CTA and noncontrast CT CAC scores was excellent (Pearson correlation coefficient = 0.96 [95% CI: 0.95, 0.97], $P < .001$; $r^2 = 0.92$, $P < .001$), as shown in Figure 5. The ICC was 0.94 (95% CI: 0.92, 0.95). The Bland-Altman plot in Figure 6 shows good accordance as well. The 95% limits of agreement were represented with the following equation: difference = $\pm 1.96 \cdot (\pi/2)^{0.5} \cdot (b + a \cdot \text{mean}^{0.5})$, where a and b were 11.4 and -13.3 , respectively.

Agreement of Risk Categorization

Agreement of risk categorization based on deep learning CTA and noncontrast CT CAC scores was excellent in the test set (weighted $\kappa = 0.94$; 95% CI: 0.91, 0.97). More specifically, 223 of 240 participants (93%) were placed into the same category according to both deep learning CTA and noncontrast CT, as shown in Tables 3 and 4. Those not placed within the same category were shifted into neighbor-

Table 3: Risk Categorization Based on DL CTA CAC Score and Noncontrast CT CAC Score in Test Set ($n = 240$)

Noncontrast CT CAC Score	DL CTA CAC Score				Total
	0–10	11–100	101–400	>400	
0–10	115*	1	0	0	116
11–100	1	50*	5	0	56
101–400	0	3	26*	3	32
>400	0	0	4	32*	36

Note.—Data are numbers of patients. CAC = coronary artery calcium, CTA = CT angiography, DL = deep learning.

* Indicates agreement of both methods.

Table 4: DL CAC Scoring from CTA Risk Categorization Agreement and Disagreement Compared with Agatston CAC Score at Noncontrast CT in Test Set ($n = 240$)

DL CTA vs Noncontrast CT Score	DL CTA CAC Score Risk Category				Total
	0–10	11–100	101–400	>400	
DL CTA higher	0	1	5	3	9
DL CTA lower	1	3	4	0	8
Same category	115*	50*	26*	32*	223*
Total	116	54	35	35	240

Note.—Data are numbers of patients. CAC = coronary artery calcium, CTA = CT angiography, DL = deep learning.

* Indicates agreement of both methods.

ing categories. Importantly, the difference between the two risk categories was never more than one.

Performance Using Different Scanner, Sex, and Section Thickness

To study the proposed method's performance using factors such as different scanner, sex, and section thickness, the differences between the noncontrast CT CAC score and the deep learning CTA CAC score were analyzed using analysis of variance. The differences between noncontrast CT and deep learning CTA CAC scores according to different factors are shown in Figure 7. There was no significant difference across the three factors (scanner [$P = .15$], sex [$P = .051$], and section thickness [$P = .67$]).

Discussion

Quantifying coronary artery calcium (CAC) scores directly from CT angiography (CTA) would eliminate the additional radiation produced at separate noncontrast CT but remains challenging. In this study, a deep learning approach was proposed to automatically quantify CAC scores at CTA. This model was developed using spectral CT, which avoided manual annotation of voxel-level calcifications. Although virtual noncontrast (VNC) CT was used to develop the algorithm, the final algorithm was tested on contrast-enhanced coronary CTA scans without spectral postprocessing. The model was developed using 365 patients

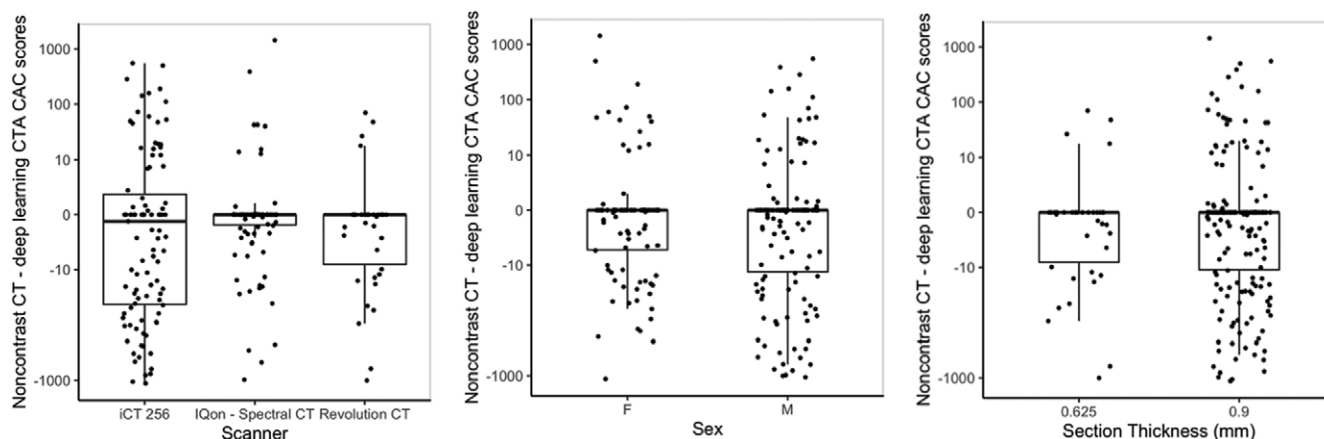


Figure 7: Box plots show difference between noncontrast CT coronary artery calcium (CAC) score and deep learning CT angiography (CTA) CAC score in test set for factors of scanner, sex, and section thickness. Lower and upper bounds of box indicate 25th and 75th percentiles. Horizontal bolded line inside box indicates median. Whiskers indicate maximums of 1.5 times the interquartile range from lower or upper quartile.

and evaluated on an independent testing set of 240 patients. The intraclass correlation coefficient (ICC) was 0.96 (95% CI: 0.95, 0.97) and 1 (95% CI: 1, 1) for the inter- and intraobserver variability at noncontrast CT semiautomatic reading, respectively. Both showed great consistency with slightly better consistency for intraobserver readings. The ICC was 1 (95% CI: 1, 1) for both the inter- and intraobserver variability of the VNC semiautomatic reading with a root mean square error of 7 or less, showing nearly perfect consistency. The excellent positive correlation (Pearson correlation coefficient = 0.95 [95% CI: 0.93, 0.97], $P < .001$; $r^2 = 0.90$, $P < .001$) was observed between semiautomatic CAC scoring obtained at noncontrast CT and VNC CT. More importantly, validation on 240 independent CTA scans from three different scanners showed excellent positive correlation (Pearson correlation coefficient = 0.96 [95% CI: 0.95, 0.97], $P < .001$; $r^2 = 0.92$, $P < .001$) with the reference Agatston score obtained at noncontrast CT. The ICC was 0.94 (95% CI: 0.92, 0.95). The agreement of risk categorization according to deep learning CTA and noncontrast CT CAC scores was excellent in the test set (weighted $\kappa = 0.94$; 95% CI: 0.91, 0.97). The resulting risk categorization was correct for 223 of 240 patients (93%). The difference between the deep learning CTA and the semiautomatic noncontrast CT CAC risk categories was never more than 1. The difference from the reference Agatston score was also robust, showing no statistical difference ($P > .05$) with regard to scanner, sex, and section thickness.

Previous attempts to quantify CAC directly from CTA usually suffered from one or all of the following three shortcomings: (a) manual or semiautomatic heuristic methods to select calcium regions (14–17), (b) using an oversimplified linear regression model to directly map calcium region volumes to Agatston scores (14–16), and (c) limited validations using images from the same scanner and population as the training set (14,15,17).

The proposed method was tailored to overcome these shortcomings. First, the calcium detection module used deep learning, which is known to handle larger appearance variations better than conventional methods, such as thresholding (27,28), to automatically detect calcium regions accurately. Second, the CAC score estimation module used deep learning to model

the complex mapping from variable CTA attenuation to CAC score instead of the oversimplified linear relationship (29). Third, scans from three different scanners, including both spectral and conventional single-energy scanners, were used to validate the proposed method, even though it was only derived from images of one spectral scanner.

Thresholding at VNC CT and keeping only components in proximity to coronary artery segmentation at CTA greatly alleviate human annotation effort in obtaining the reference calcium region mask during the training phase (18) by taking advantage of perfectly aligned CTA and VNC scans obtained with the spectral CT scanner. VNC CT attenuation was close enough to real noncontrast CT (30) such that VNC CT could be used to estimate the CAC score (31,32), which is also confirmed in this study by the high correlation between VNC CT and noncontrast CT CAC scores (Pearson correlation coefficient = 0.95 [95% CI: 0.93, 0.97], $P < .001$) in the spectral subset of the test set. Using the CAC segmentation mask as an intermediate task enabled the calcium detection module to learn effectively from detailed voxel-level supervision. It also allowed the CAC score regression module to focus on CAC regions to estimate the corresponding CAC score.

Our study had some limitations. First, although the test set contains scans from three different scanners, the training set scans were obtained from a single spectral CT scanner. Second, to avoid the tedious and error-prone manual annotation and registration, VNC CT instead of noncontrast CT was used to obtain the calcium regions during the training phase. Third, although the test set is temporally independent from the training and validation set and from different scanners, all images were from the same hospital in this study. As a next step, we plan to collect additional images from more centers to further evaluate performance in the near future.

To conclude, a deep learning method was proposed to automatically quantify the coronary artery calcium (CAC) score from CT angiography (CTA) alone. The results demonstrate that the proposed method could accurately quantify CAC as well as categorize risk at CTA.

Acknowledgments: The authors would also like to thank Xingbiao Chen, PhD; Ya Pei, BS; Jubin Zhao, BS; Qiming Shi, BS; and Yongzhen Cui, BS, for their efforts in data collection and processing.

Author contributions: Guarantors of integrity of entire study, J.L., K.H., U.J.S., B.Z.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, D.M., J.B., J.L., H.Y.Y., Y.Y., U.J.S., B.Z.; clinical studies, D.M., J.B., J.L., H.Y.Y., J.Z.; statistical analysis, D.M., J.B., J.L., Z.Q., H.Y.Y., J.Z.; and manuscript editing, D.M., J.B., W.C., H.Y., J.L., K.Y., H.L., K.H., H.Y.Y., J.Z., Y.Y., H.W.M., U.J.S., B.Z.

Disclosures of Conflicts of Interest: D.M. No relevant relationships. J.B. Provisional patent with Keya Medical. W.C. No relevant relationships. H.Y. No relevant relationships. J.L. No relevant relationships. K.Y. No relevant relationships. H.L. No relevant relationships. Z.Q. No relevant relationships. K.H. No relevant relationships. H.Y.Y. No relevant relationships. J.Z. No relevant relationships. Y.Y. No relevant relationships. H.W.M. No relevant relationships. U.J.S. Grants from Astellas, Bayer, Bracco, Elucid BioImaging, GE, Guerbet, HeartFlow, and Siemens; consulting fees from Elucid BioImaging, GE, and Guerbet; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Guerbet, HeartFlow, and Siemens; support for attending meetings and/or travel from Astellas, Bayer, GE, Guerbet, and Siemens. B.Z. No relevant relationships.

References

- Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson KE. Coronary Calcium Score and Cardiovascular Risk. *J Am Coll Cardiol* 2018;72(4):434–447.
- Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 1990;15(4):827–832.
- Wang W, Wang H, Chen Q, et al. Coronary artery calcium score quantification using a deep-learning algorithm. *Clin Radiol* 2020;75(3):237.e11–237.e16.
- Yu M, Shen C, Dai X, et al. Clinical Outcomes of Dynamic Computed Tomography Myocardial Perfusion Imaging Combined With Coronary Computed Tomography Angiography Versus Coronary Computed Tomography Angiography-Guided Strategy. *Circ Cardiovasc Imaging* 2020;13(1):e009775.
- Zhang N, Yang G, Zhang W, et al. Fully automatic framework for comprehensive coronary artery calcium scores analysis on non-contrast cardiac-gated CT scan: Total and vessel-specific quantifications. *Eur J Radiol* 2021;134:109420.
- Işgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging* 2012;31(12):2322–2334.
- de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Işgum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans Med Imaging* 2019;38(9):2127–2138.
- Shadmi R, Mazo V, Bregman-Amitai O, Elnekave E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, April 4–7, 2018. Piscataway, NJ: IEEE, 2018; 24–28.
- Takx RAP, de Jong PA, Leiner T, et al. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLoS One* 2014;9(3):e91239.
- van Velzen SGM, Lessmann N, Velthuis BK, et al. Deep learning for automatic calcium scoring in CT: Validation using multiple cardiac CT and chest CT protocols. *Radiology* 2020;295(1):66–79.
- Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ. Automated Agatston Score Computation in non-ECG Gated CT Scans Using Deep Learning. *Proc SPIE Int Soc Opt Eng* 2018;10574:105742K.
- Miller JM, Rochitte CE, Dewey M, et al. Diagnostic performance of coronary angiography by 64-row CT. *N Engl J Med* 2008;359(22):2324–2336.
- Gao Z, Wang X, Sun S, et al. Learning physical properties in complex visual scenes: An intelligent machine for perceiving blood flow dynamics from static CT angiography imaging. *Neural Netw* 2020;123:82–93.
- Mylonas I, Alam M, Amily N, et al. Quantifying coronary artery calcification from a contrast-enhanced cardiac computed tomography angiography study. *Eur Heart J Cardiovasc Imaging* 2014;15(2):210–215.
- Schuhbaeck A, Otaki Y, Achenbach S, et al. Coronary calcium scoring from contrast coronary CT angiography using a semiautomated standardized method. *J Cardiovasc Comput Tomogr* 2015;9(5):446–453.
- Ahmed W, de Graaf MA, Broersen A, et al. Automatic detection and quantification of the Agatston coronary artery calcium score on contrast computed tomography angiography. *Int J Cardiovasc Imaging* 2015;31(1):151–161.
- Teßmann M, Vega-Higuera F, Bischoff B, Hausleiter J, Greiner G. Automatic detection and quantification of coronary calcium on 3D CT angiography data. *Comput Sci Res Dev* 2011;26(1–2):117–124.
- Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Işgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal* 2016;34:123–136.
- Danad I, Fayad ZA, Willemink MJ, Min JK. New applications of cardiac computed tomography: Dual-energy, spectral, and molecular CT imaging. *JACC Cardiovasc Imaging* 2015;8(6):710–723.
- Numburi UD, Schoenhagen P, Flamm SD, et al. Feasibility of dual-energy CT in the arterial phase: Imaging after endovascular aortic repair. *Am J Roentgenol* 2010;195(2):486–493.
- Kaufmann S, Sauter A, Spira D, et al. Tin-filter enhanced dual-energy-CT: image quality and accuracy of CT numbers in virtual noncontrast imaging. *Acad Radiol* 2013;20(5):596–603.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Cham, Switzerland: Springer, 2015; 234–241.
- Kong B, Wang X, Bai J, et al. Learning tree-structured representation for 3D coronary artery segmentation. *Comput Med Imaging Graph* 2020;80:101688.
- Guo Z, Bai J, Lu Y, et al. DeepCenterline: A Multi-task Fully Convolutional Network for Centerline Extraction. In: *Information Processing in Medical Imaging*. Springer Verlag, 2019; 441–453.
- Wu D, Wang X, Bai J, et al. Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs. *Int J Comput Assist Radiol Surg* 2019;14(2):271–280.
- Sevrakov AB, Bland JM, Kondos GT. Serial electron beam CT measurements of coronary artery calcium: Has your patient's calcium score actually changed? *AJR Am J Roentgenol* 2005;185(6):1546–1553.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–444.
- He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Computer Vision and Pattern Recognition*, 2016; 770–778.
- Leshno M, Schocken S. Multilayer Feedforward Networks with a Non-Polynomial Activation Functions Can Approximate Any Function. *Neural Netw* 1993;6(6):861–867.
- Toepker M, Moritz T, Krauss B, et al. Virtual non-contrast in second-generation, dual-energy computed tomography: reliability of attenuation values. *Eur J Radiol* 2012;81(3):e398–e405.
- Gassert FG, Schacky CE, Müller-Leisse C, et al. Calcium scoring using virtual non-contrast images from a dual-layer spectral detector CT: comparison to true non-contrast data and evaluation of proportionality factor in a large patient collective. *Eur Radiol* 2021;31(8):6193–6199.
- Yamada Y, Jinzaki M, Okamura T, et al. Feasibility of coronary artery calcium scoring on virtual unenhanced images derived from single-source fast kVp-switching dual-energy coronary CT angiography. *J Cardiovasc Comput Tomogr* 2014;8(5):391–400.