Korean Journal of Radiology

KJR

Fully Automatic Coronary Calcium Score Software Empowered by Artificial Intelligence Technology: Validation Study Using Three CT Cohorts

June-Goo Lee[1], HeeSoo Kim[2], Heejun Kang[3], Hyun Jung Koo[2], Joon-Won Kang[2], Young-Hak Kim[3]*, Dong Hyun Yang[2]*

[1]Biomedical Engineering Research Center, Asan Institute for Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; [2]Department of Radiology and Research Institute of Radiology, Cardiac Imaging Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; [3]Divison of Cardiology, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Objective: This study aimed to validate a deep learning-based fully automatic calcium scoring (coronary artery calcium [CAC]_auto) system using previously published cardiac computed tomography (CT) cohort data with the manually segmented coronary calcium scoring (CAC_hand) system as the reference standard.

Materials and Methods: We developed the CAC_auto system using 100 co-registered, non-enhanced and contrast-enhanced CT scans. For the validation of the CAC_auto system, three previously published CT cohorts (n = 2985) were chosen to represent different clinical scenarios (i.e., 2647 asymptomatic, 220 symptomatic, 118 valve disease) and four CT models. The performance of the CAC_auto system in detecting coronary calcium was determined. The reliability of the system in measuring the Agatston score as compared with CAC_hand was also evaluated per vessel and per patient using intraclass correlation coefficients (ICCs) and Bland-Altman analysis. The agreement between CAC_auto and CAC_hand based on the cardiovascular risk stratification categories (Agatston score: 0, 1–10, 11–100, 101–400, > 400) was evaluated.

Results: In 2985 patients, 6218 coronary calcium lesions were identified using CAC_hand. The per-lesion sensitivity and false-positive rate of the CAC_auto system in detecting coronary calcium were 93.3% (5800 of 6218) and 0.11 false-positive lesions per patient, respectively. The CAC_auto system, in measuring the Agatston score, yielded ICCs of 0.99 for all the vessels (left main 0.91, left anterior descending 0.99, left circumflex 0.96, right coronary 0.99). The limits of agreement between CAC_auto and CAC_hand were 1.6 ± 52.2. The linearly weighted kappa value for the Agatston score categorization was 0.94. The main causes of false-positive results were image noise (29.1%, 97/333 lesions), aortic wall calcification (25.5%, 85/333 lesions), and pericardial calcification (24.3%, 81/333 lesions).

Conclusion: The atlas-based CAC_auto empowered by deep learning provided accurate calcium score measurement as compared with manual method and risk category classification, which could potentially streamline CAC imaging workflows.

Keywords: *Coronary artery calcium score; Computed tomography; Artificial intelligence; Accuracy*

INTRODUCTION

Since Arthur Agatston and his radiologist colleague, Warren Janowitz, proposed the coronary artery calcium (CAC) score based on electron beam computed tomography (CT) in 1990 [1], the Agatston score has been used to assess the risk of cardiovascular clinical outcomes [2,3]. More than 30 years after its introduction, CAC scoring still requires manual inputs from skilled professionals, which is associated with prolonged processing time [4,5]. Before the era of deep learning, researchers proposed feature-based or atlas-based models (as well as combined feature- and atlas-based models) for CAC scoring automation [6-10]. Some researchers have recently proposed a deep learning-based automatic CAC scoring (CAC_auto) system, with the reliability of calcium measurement mainly being evaluated using per-patient comparison [11-14] or per-vessel comparison [15]. Detailed error analysis is essential for applying any automatic algorithm to routine imaging workflows, and, for this, per-vessel and per-lesion analyses may be more suitable than per-patient analyses. Additionally, analyses of the causes of errors can help improve the algorithms for CAC_auto.

For the training of artificial intelligence (AI) algorithms based on the end-to-end training strategy [10-12], manual labels for the location of CAC (e.g., right coronary artery [RCA] or left main artery [LM]) are required. A weakness of this end-to-end training strategy may be that calcium near the coronary artery (e.g., cardiac valves or aortic wall calcium) can be mistaken for CAC. Moreover, the slice-by-slice inference of CAC may take a relatively long time [1]. If an algorithm can identify the coronary tree, aortic root, and cardiac ventricles automatically in a CAC CT image without intravenous contrast enhancement, CAC_auto can be performed easily by including only calcium that overlaps the coronary artery tree region.

We developed a deep learning-based fully automatic calcium scoring system using non-enhanced CT, the atlas-based CAC_auto. This study aimed to validate the atlas-based CAC_auto using previously published cardiac CT cohort data from a single institution with the manually segmented CAC scoring (CAC_hand) system as the reference standard.

MATERIALS AND METHODS

The Institutional Review Board of Asan Medical Center (Seoul, Korea) approved this study, and the need for informed consent was waived due to its retrospective design (IRB No. 2018-1155). To minimize the possibility of unintentionally sharing information that can be used to re-identify private information, a subset of the data generated for this study are available at the center for open science websites (https://osf.io/mu5k8/).

Development of CAC_Auto

The concept of the atlas-based CAC_auto, empowered by deep learning technology, is illustrated in Figure 1. If we know the spatial information (e.g., coronary artery, aortic valve, left ventricular myocardium) of a CAC image, CAC_auto can be performed easily. To perform semantic segmentation on a non-enhanced CAC image, we required a 'labeled dataset' for the training of a segmentation algorithm. However, the labeling of coronary and non-coronary structures on a non-enhanced CAC image can be challenging because of the low attenuation differences between lesions. Therefore, we first generated a labeled mask on coronary CT angiography (CCTA) (Step 1 in Fig. 1). For the development dataset, 100 CCTA images (n = 100, mean age 64.1 ± 8.2 years; 76.3% male; mean body mass index, 25.3 ± 2.3 kg/m^2; three CT vendors and six scanners) were chosen from a cohort used in a previously published investigation of CCTA-based myocardial segmentation [16]. All labeled masks in the development dataset were confirmed by a cardiac radiologist. The second step was the transfer of spatial information from CCTA to CAC images for obtaining the 'labeled CAC dataset' (Step 2 in Fig. 1) by image registration of the CCTA and CAC. The registration process, consisting of affine registration followed by B-spline registration, performed using the elastix [17,18], was applied to align the CAC images to the CCTA images with the labeled masks. This allowed a voxel-by-voxel comparison of the registered images. The final step was to develop a deep learning algorithm for the semantic segmentation of CAC images (Step 3 in Fig. 1). The basic network architecture is based on the three-dimensional (3D) U-Net [19], which was designed to input the 3D patch extracted from volumetric CT images and output the 3D patch from the multi-label semantic images. The transferred coronary artery region mask was dilated twice with 2 x 2 x 1 morphological structuring elements to compensate for misregistration errors and include nearby regions. It interpolated the 3D CT images to have 1 mm resolutions in the x-, y-, and z-directions. The intensity of the voxels was adjusted with window width and level values of 350
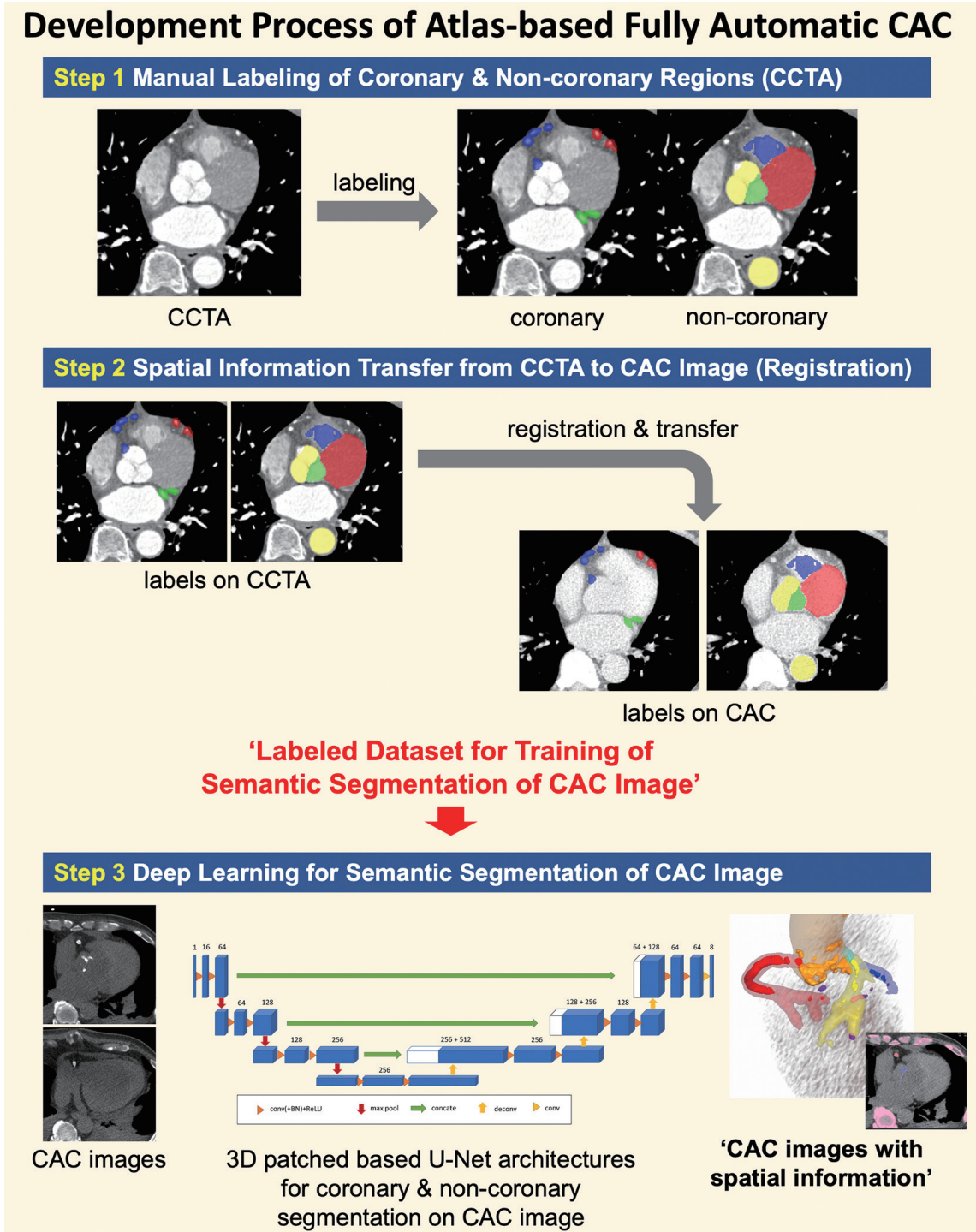
Fig. 1. Development process of the atlas-based coronary calcium scoring software. Step 1. Manual labeling of coronary and non-coronary regions using CCTA. Step 2. Spatial information transfer from CCTA to CAC images without contrast enhancement using image registration. Step 3. Deep learning for semantic segmentation of CAC images. CAC = coronary artery calcium, CCTA = coronary computed tomography angiography, 3D = three-dimensional

and -100 Houndsfield units (HU), respectively. The network input consisted of 3D-patch-based CT images with 64 × 64 × 64 voxels for each, which were randomly sampled for training procedures, similar to the structure used in our previous study [20]. The loss function of the network was the cross-entropy loss function, and the optimizer used was RMSprop. The training was performed with eight mini-batches of the 3D patch, a 10^{-5} learning rate, and 105

iterations. The two 3D-patch-based U-Net models were trained separately for coronary artery regions and other surrounding structures, including the myocardium, left and right ventricle, aorta, papillary muscles, and rib. Calcium lesions were detected when the initial candidate was in contact with the coronary artery region, and it did not belong to other structures.

Validation Dataset

Three previously published CCTA cohort datasets were chosen for the CAC_auto validation dataset as follows: 1)

asymptomatic population for health check-up (screening group; n = 2653) [21]; 2) symptomatic patients who underwent invasive fractional flow reserve measurements (FFR group; n = 222) [22]; and 3) patients with mitral valve prolapse who underwent preoperative CCTA (valve group, n = 145) [23]. Among the 3020 individuals in these cohorts, 35 were excluded because of the lack of CAC image data, previous coronary stent insertion, or overlapping of cohorts. Finally, 2985 individuals were included in the validation dataset. The baseline and CT imaging characteristics are presented in Table 1. The screening, FFR, and valve groups

Table 1. Characteristics of Validation Datasets

Characteristic	Screening Group (n = 2647)	FFR Group (n = 220)	Valve Group (n = 118)	All (n = 2985)
Baseline characteristics				
Study period	January 2008–December 2009	May 2011–July 2015	May 2011–October 2013	
Age, years	54.1 ± 9.2	60.9 ± 10.3	52.7 ± 14.1	54.5 ± 9.7
Male	1807 (68.3)	169 (76.8)	81 (68.6)	2057 (68.9)
Body mass index, kg/m²	24.9 ± 3.3	25.2 ± 2.8	24.1 ± 3.1	24.8 ± 3.3
Diabetes mellitus	385 (14.5)	50 (22.7)	9 (7.6)	444 (14.9)
Hypertension*	908 (34.3)	128 (58.2)	38 (32.2)	1074 (36.0)
Current smoker	614 (23.2)	65 (29.5)	8 (6.8)	687 (23.0)
Hyperlipidemia†	422 (15.9)	153 (69.5)	66 (55.9)	641 (21.5)
Agatston calcium score‡	0.0 (0.0–18.5)	76.1 (10.3–256.2)	0.0 (0.0–25.3)	0.0 (0.0–29.3)
Agatston calcium score risk category				
0	1654 (62.5)	43 (19.5)	64 (54.2)	1761 (59.0)
1–10	219 (8.3)	12 (5.5)	18 (15.3)	249 (8.3)
11–100	491 (18.5)	67 (30.5)	24 (20.3)	582 (19.5)
101–400	205 (7.7)	66 (30.0)	6 (5.1)	277 (9.3)
> 400	78 (2.9)	32 (14.5)	6 (5.1)	116 (3.9)
CT imaging characteristics				
Scanner type				
GE discovery 750 HD	0 (0.0)	18 (8.2)	1 (0.8)	19 (0.6)
GE lightspeed VCT	2353 (88.9)	13 (5.9)	0 (0.0)	2366 (79.3)
Siemens definition	294 (11.1)	94 (42.7)	41 (34.8)	429 (14.4)
Siemens definition flash	0 (0.0)	95 (43.2)	76 (64.4)	171 (5.7)
ECG triggering	Yes	Yes	Yes	Yes
Tube voltage, kVp				
100	0 (0.0)	15 (6.8)	0 (0.0)	15 (0.5)
120	2647 (100.0)	205 (93.2)	118 (100.0)	2970 (99.5)
Slice thickness, mm				
2.5	2353 (88.9)	31 (14.1)	1 (0.8)	2385 (79.9)
3.0	294 (11.1)	189 (85.9)	117 (99.2)	600 (20.1)
Reconstruction kernel				
B30f	166 (6.3)	5 (2.3)	4 (3.4)	175 (5.9)
B35f	128 (4.8)	184 (83.6)	113 (95.8)	425 (14.2)
STANDARD	2353 (88.9)	31 (14.1)	1 (0.8)	2385 (79.9)

Unless specified otherwise, continuous variables are presented as mean ± standard deviation, and categorical variables are expressed number (percentage). *Defined as receiving antihypertensive treatment or having systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg, †Defined as total cholesterol > 200 mg/dL or receiving anti-lipidemic treatment, ‡Agatston calcius score are presented as median and interquartile range due to the skewed distribution of this variable. FFR = fractional flow reserve

underwent CT examinations between 2008 and 2009, 2011 and 2015, and 2011 and 2013, respectively. Among the 2985 enrollees, 1761 (59.0%), 249 (8.3%), 582 (19.5%), 277 (9.3%), and 116 (3.9%) belonged to the CAC 0, 1–10, 11–100, 101–400, and > 400 risk categories, respectively. Four types of CT machines from two companies (GE, Siemens) were used in this study, and all the scans were taken in electrocardiography-triggered mode with a slice thickness of 2.5 or 3 mm.

Reference Standard Calcium Scoring: CAC_Hand

CAC_hand was performed by an experienced radiology technologist (> 500 CAC examinations) in all cases. An expert cardiac radiologist (15 years of experience, > 50000 CAC and CCTA examinations) reviewed all the labeled results by the technologist. CAC_hand was performed using a research prototype (AVIEW CAC, Coreline Soft, Co. Ltd.) without the assistance of an automatic algorithm. After thresholding (> 130 HU) with a colored overlay, the observer manually labeled coronary calcifications according to their anatomical locations (i.e., LM, left anterior descending artery [LAD], left circumflex artery [LCX], RCA arteries). The labeled mask containing binary information indicating the presence or absence of calcium in each image slice was saved for per-lesion analysis.

Automatic Calcium Scoring: CAC_Auto

CAC_auto can be performed via batch processing without user interactions. After all CT data were loaded in a desktop computer (Intel Core 17-8700 3.19 GHz, 32 GB RAM, NVIDIA GeForce GTX 1060 6 GB), a research prototype software (AVIEW CAC, Coreline Soft, Co. Ltd.) automatically determined the CAC score, and there was no need to open the CT images (Fig. 2). The processing time for CAC_auto was recorded, and all the CAC results were downloaded as a single comma-separated value file. The labeled mask for the CAC_auto was saved for per-lesion comparison with the CAC_hand mask.

Per-Lesion Comparisons

Per-lesion comparisons between CAC_auto and CAC_hand were performed using the binary image mask described elsewhere [9]. For each CAC scan, the initial masks were generated by connected component analysis after applying thresholding at 130 HU and discarding all masks that were less than 2.5 mm³. Candidate coronary calcium masks were selected based on the coronary region and other structural information generated by CAC_auto. The masks
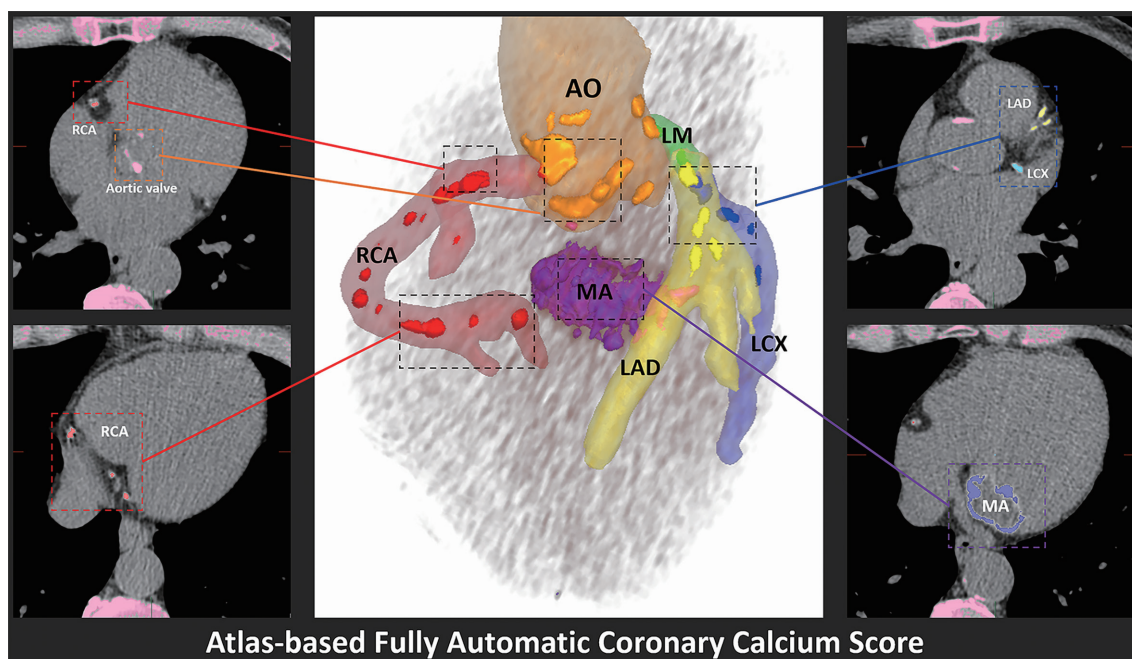
Fig. 2. CAC_auto result for a patient with severe calcifications in both the aortic and mitral valves. CAC_auto can differentiate the coronary artery tree mask, aortic root mask (AO), and other cardiac area masks for non-enhanced calcium scoring of computed tomography images (central image with color masks). It is possible to distinguish calcium located in the heart region from that in the coronary artery or other areas, such as the aortic valve and MA. Additionally, the coronary tree mask enables the localization of calcium within blood vessels (i.e., LM, LAD, LCX, and RCA). AO = aorta, CAC_auto = automatic coronary artery calcium scoring, LAD = left anterior descending artery, LCX = left circumflex artery, LM = left main artery, MA = mitral annulus, RCA = right coronary artery

were compared with manual masks. The overlapping lesions for the CAC_auto and CAC_hand were defined as coronary calcium, and the mismatched lesions were divided into false-positive results (e.g., valve or aortic wall calcifications) and false-negative results (i.e., coronary calcifications missed by CAC_auto) errors. The per-lesion sensitivity was calculated as (true positive lesion number/ total lesion number) x 100. The false-positive rate was calculated as (number of false-positive lesions/total patient number). All mismatched lesions were subsequently reviewed by a senior cardiovascular radiologist to analyze the causes of mismatches and lesion locations.

Statistical Analyses

The Agatston score and calcium volume (mm^3) were calculated using CAC_auto and CAC_hand, and the values were compared. The reliability of the Agatston score and volume calculation using CAC_auto as compared with CAC_hand was evaluated using the intraclass correlation coefficient (ICC). Bland-Altman analysis was used to calculate the 95% limits of agreement between CAC_auto and CAC_hand. For cardiovascular risk stratification using the Agatston score, the following categories were used: 0,

1–10, 11–100, 101–400, and < 400. The risk stratification results for CAC_auto and CAC_hand were compared using Cohen's linearly weighted kappa statistic.

RESULTS

CAC_Auto for Coronary Calcium Measurement

For the Agatston score, CAC_auto yielded high ICCs (0.99) for all the patients in all the cohorts combined, as well as the screening, FFR, and valve groups, individually (Table 2). For the per-vessel evaluation, the ICCs for the Agatston scores of the LM, LAD, LCX, and RCA were 0.91, 0.99, 0.96, and 0.99, respectively (Fig. 3, Supplementary Fig. 1). In the Bland-Altman plot analysis, the 95% limit of agreement for the Agatston score was 1.6 ± 52.2, when all the cohorts were combined (Fig. 4). Systemic deviations did not occur in any subgroup or during the per-vessel analysis. The limits of agreement tended to increase in the severe CAC group compared with the mild CAC group (CAC 0–100, -0.4 ± 15.5; CAC 11–100, 5.5 ± 93.5; CAC > 400, 37.7 ± 197.0). The reliability of the calcium volume (mm^3) was also high, with similar values yielded by the Agatston score measurement for both the per-patient and per-vessel analyses (Table 2). The

Table 2. CAC_Auto for Calcium Volume and Agatston Calcium Score Measurements

Index	Total	LM	LAD	LCX	RCA
All					
Agatston score, ICC	0.99 (0.99, 0.99)	0.91 (0.90, 0.91)	0.99 (0.99, 0.99)	0.96 (0.96, 0.96)	0.99 (0.99, 0.99)
Agatston score, limits of agreement	1.6 ± 52.2	-1.1 ± 28.3	2.1 ± 37.4	0.3 ± 30.2	0.3 ± 23.8
Volume, mm^3, ICC	0.99 (0.99, 0.99)	0.91 (0.90, 0.92)	0.99 (0.99, 0.99)	0.96 (0.96, 0.96)	0.99 (0.99, 0.99)
Volume, mm^3, limits of agreement	1.1 ± 42.1	-0.9 ± 22.9	1.6 ± 28.9	0.2 ± 24.8	0.2 ± 20.1
Screening group					
Agatston score, ICC	0.99 (0.99, 0.99)	0.91 (0.90, 0.91)	0.98 (0.98, 0.98)	0.95 (0.95, 0.95)	0.99 (0.99, 0.99)
Agatston score, limits of agreement	2.1 ± 52.4	-0.6 ± 26.8	1.8 ± 37.1	0.3 ± 29.4	0.6 ± 22.7
Volume, mm^3, ICC	0.99 (0.99, 0.99)	0.91 (0.90, 0.92)	0.98 (0.98, 0.98)	0.95 (0.94, 0.95)	0.99 (0.99, 0.99)
Volume, mm^3, limits of agreement	1.5 ± 41.7	-0.5 ± 21.7	1.4 ± 28.5	0.2 ± 24.2	0.4 ± 18.4
FFR group					
Agatston score, ICC	0.99 (0.99, 0.99)	0.89 (0.84, 0.92)	0.99 (0.99, 0.99)	0.98 (0.98, 0.99)	0.99 (0.99, 0.99)
Agatston score, limits of agreement	0.3 ± 41.3	-5.9 ± 41.0	5.8 ± 44.2	1.7 ± 40.8	-1.3 ± 25.5
Volume, mm^3, ICC	0.99 (0.99, 0.99)	0.90 (0.86, 0.92)	0.99 (0.99, 0.99)	0.98 (0.98, 0.99)	0.99 (0.99, 0.99)
Volume, mm^3, limits of agreement	-0.1 ± 33.2	-4.8 ± 33.1	4.3 ± 35.0	1.4 ± 33.0	-1.1 ± 20.8
Valve group					
Agatston score, ICC	0.99 (0.99, 0.99)	0.93 (0.91, 0.95)	0.99 (0.99, 0.99)	0.91 (0.88, 0.94)	0.97 (0.96, 0.98)
Agatston score, limits of agreement	-6.4 ± 61.9	-2.6 ± 29.7	0.8 ± 27.8	-1.7 ± 23.5	-2.9 ± 37.9
Volume, mm^3, ICC	0.99 (0.98, 0.99)	0.94 (0.92, 0.96)	0.99 (0.99, 0.99)	0.91 (0.87, 0.94)	0.95 (0.93, 0.96)
Volume, mm^3, limits of agreement	-6.2 ± 60.7	-2 ± 23.5	0.5 ± 21.6	-1.6 ± 19.5	-3.1 ± 42.0

Numbers in parentheses indicate 95% confidence intervals. CAC_auto = automatic coronary artery calcium scoring, ICC = intraclass correlation coefficient, LAD = left anterior descending artery, LCX = left circumflex artery, LM = left main artery, RCA = right coronary artery
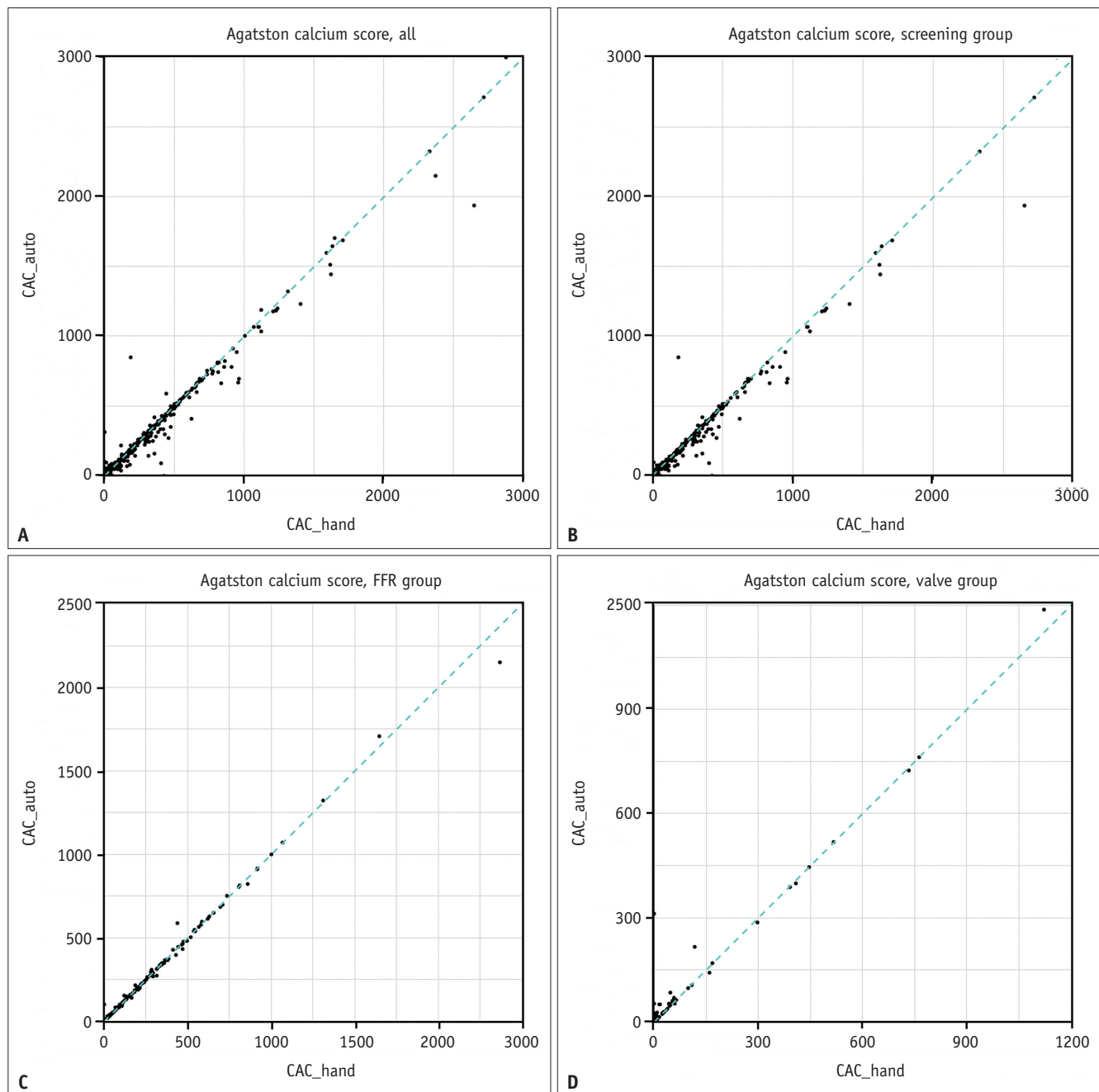
Fig. 3. Scatter plots showing the correlation between CAC_hand and CAC_auto.
A. All. **B.** Screening group. **C.** FFR group. **D.** Valvular heart disease (valve) group. CAC_auto = automatic coronary artery calcium scoring, CAC_hand = manually segmented coronary calcium scoring, FFR = fractional flow reserve

mean CAC_auto computation duration ± standard deviation was 57 ± 11 seconds (range, 41–80 seconds).

CAC_Auto for Risk Category by Agatston Score Assignment

CAC_auto achieved high weighted kappa values of 0.94, 0.94, 0.96, and 0.82 for all patients and the screening, FFR, and valve groups, respectively (Table 3). The CAC_

auto assigned the participants to the correct cardiovascular risk category 93.9% of the time (2804/2985) and to a neighboring category 5.3% of the time (159/2985). Only 0.7% (22/2985) of participants had more than two category differences between CAC_auto and CAC_hand (Fig. 2, Table 3).

Per-Lesion Analysis

Among 2985 enrollees, 6218 lesions were identified using
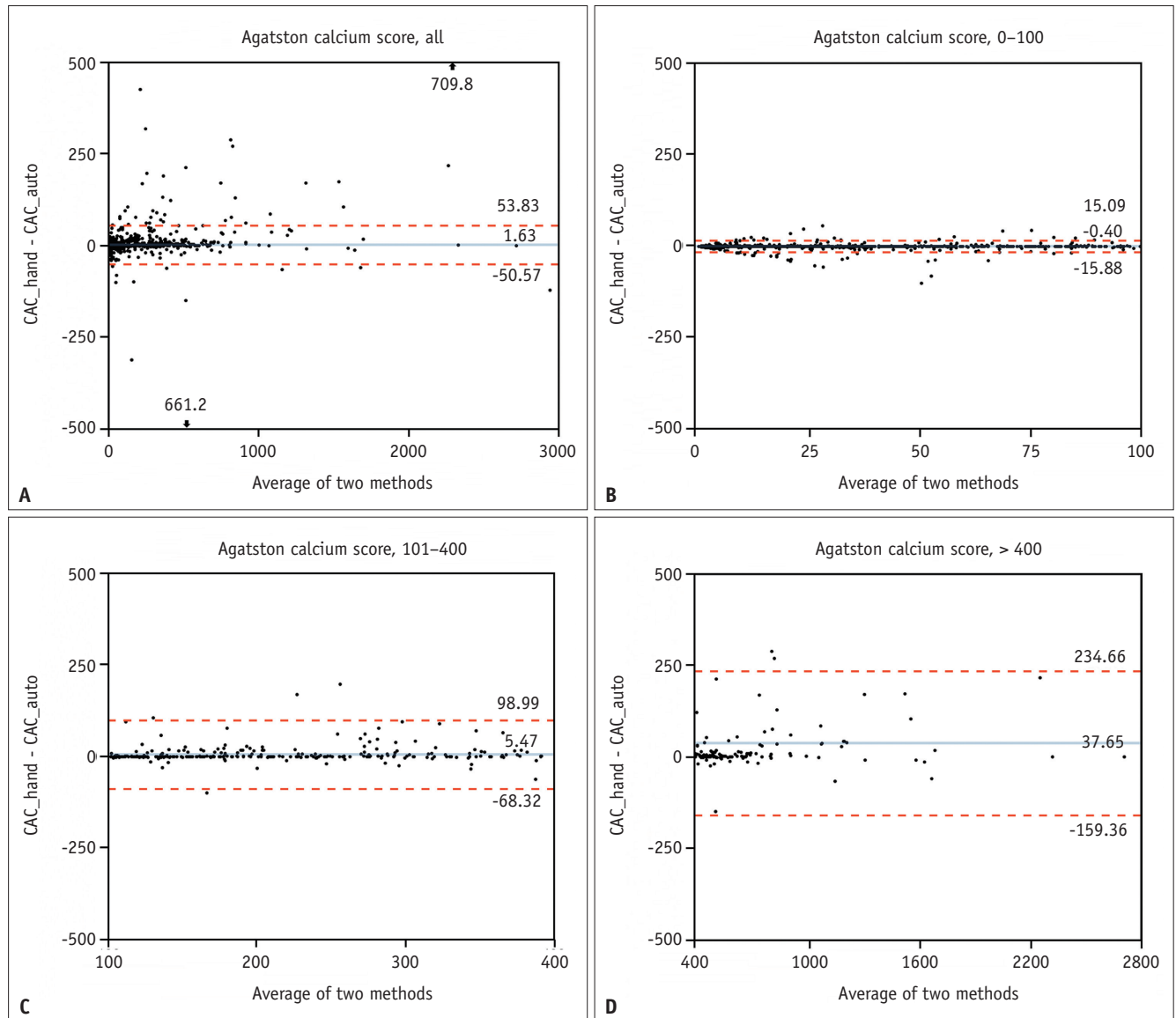
Fig. 4. Bland-Altman analysis graphs for CAC_hand CAC_auto.
A. All. **B.** CAC 0–100 group. **C.** CAC 101–400 group. **D.** CAC > 400 group. Outliers are indicated by arrows and their values. CAC_auto = automatic coronary artery calcium scoring, CAC_hand = manually segmented coronary calcium scoring

CAC_hand and compared with calcium lesions categorized by CAC_auto. CAC_auto yielded a per-lesion sensitivity of 93.3% (5800 of 6218 lesions) and a false-positive rate of 0.11 per patient (333 lesions among 2985 patients). The main causes of false positive results were image noise or artifacts (29.1%, 97/333) and aortic wall calcification (25.5%, 85/333) (Table 4, Fig. 5). Among 85 aortic wall calcifications falsely detected by CAC_auto, 64 (75.3%) were classified as LM lesions. Among 81 false-positive pericardial calcifications, 26 (32.1%) were categorized as LCX lesions. Of interest, 15.9% (53/333) of the false-positive lesions were judged to be human errors committed

during the initial labeling for the CAC_hand. In these cases, CAC_auto facilitated the detection of lesions that were neglected during human labeling due to the small size of calcium or confusion with surrounding image noise (Fig. 5). After excluding image noise (or artifacts) and human error, the number of 'anatomically' false-positive results decreased to 183 (0.06 false-positive lesions per patient). Among the 2985 patients, 161 (5.4%) showed calcification at the mitral valve or annulus, and three mitral annular calcifications were misclassified as LCX lesions. Thus, the false-positive rate of CAC_auto regarding mitral calcification was 1.9% (3/161). There were 418

Table 3. CAC_Auto for Risk Category by Agatston Score Assignment

CAC_Auto \ CAC_Hand	0 (n = 1761)	1–10 (n = 249)	11–100 (n = 582)	101–400 (n = 277)	> 400 (n = 116)	Linearly Weighted Kappa (95% CI)
All						0.94 (0.94, 0.95)
0	1650	11	5	0	1	
1–10	96	223	12	0	0	
11–100	13	15	561	11	1	
101–400	2	0	4	264	8	
> 400	0	0	0	2	106	
Screening group						0.94 (0.94, 0.95)
0	1562	10	5	0	1	
1–10	84	199	11	0	0	
11–100	8	10	472	9	1	
101–400	0	0	3	194	6	
> 400	0	0	0	2	70	
FFR group						0.96 (0.93, 0.99)
0	37	1	0	0	0	
1–10	5	11	0	0	0	
11–100	0	0	66	2	0	
101–400	1	0	1	64	1	
> 400	0	0	0	0	31	
Valve group						0.82 (0.73, 0.90)
0	51	0	0	0	0	
1–10	7	13	1	0	0	
11–100	5	5	23	0	0	
101–400	1	0	0	6	1	
> 400	0	0	0	0	5	

CAC_auto = automatic coronary artery calcium scoring, CAC_hand = manually segmented coronary calcium scoring, CI = confidence interval, FFR = fractional flow reserve, kappa = linearly weighted kappa value

**Table 4. Mismatched Lesions between CAC_Auto and CAC_Hand**

Category	Data
False-positive errors (n = 333)	
Image noise or motion artifact	97 (29.1)
Aortic wall	85 (25.5)
Pericardium	81 (24.3)
Human error	53 (15.9)
Myocardium	10 (3.0)
Mitral valve or annulus	3 (0.9)
Rib	3 (0.9)
Lung	1 (0.3)
False-negative errors (n = 418)	
LM	41 (9.8)
LAD	106 (25.4)
LCX	107 (25.6)
RCA	164 (39.2)

Data are number with percentage in the parentheses. CAC_auto = automatic coronary artery calcium scoring, CAC_hand = manually segmented coronary calcium scoring, LAD = left anterior descending artery, LCX = left circumflex artery, LM = left main artery, RCA = right coronary artery

false-negative results, and the most common location of missed lesions was the RCA (n = 164, 39.2%), followed by the LCX, LAD, and LM (Table 4).

DISCUSSION

We developed a novel atlas-based CAC_auto system and validated it using three CT angiography cohort datasets and a manually generated reference standard (CAC_hand). The major findings of this study were as follows: 1) CAC_auto demonstrated high reliability for the coronary calcium score (ICC 0.99) and volume (ICC 0.99) measurements, and it yielded a high accuracy for the assignment of risk categories (kappa value 0.94); 2) during the lesion-based analysis, CAC_auto provided a high sensitivity (93.3%) and a low false-positive rate (0.11 false-positive lesions per patient); 3) among the 333 false-positive lesions, the true anatomical false-positive rate, except for artifacts and human errors, was as low as 0.06 lesions per patient. Several of the anatomical

False Positive Calcium Lesions on CAC_auto
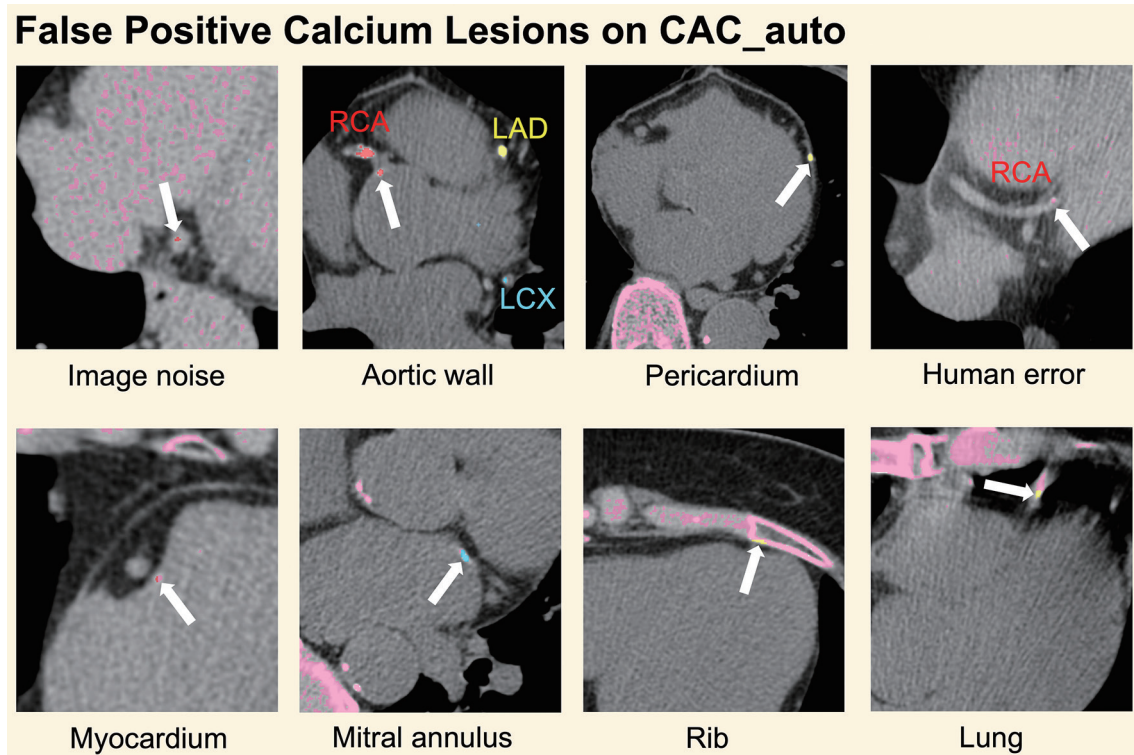
Fig. 5. Mismatched lesions detected based on CAC_hand and CAC_auto. Examples of mismatched lesions (arrows) detected based on CAC_hand and CAC_auto. In the areas (pink; spine, rib, and image noise) where the Hounsfield number was 130 or more, each coronary artery calcification was marked with a different color (i.e., LAD, yellow; LCX, blue; RCA, red). Human error was identified when CAC_hand did not lead to the detection of a small calcification in the RCA, but CAC_auto did. CAC_auto = automatic coronary artery calcium scoring, CAC_hand = manually segmented coronary calcium scoring, LAD = left anterior descending artery, LCX = left circumflex artery, RCA = right coronary artery

errors were caused by the aortic wall and pericardial calcium, and the extracardiac error rate was low.

We followed the atlas-based approach [8,9] but incorporated a deep-learning-based semantic segmentation model to replace the time-consuming non-rigid registration of the multi-atlas method. The false-positive rate of 0.11 per patient in the present study was significantly lower than what was reported previously for the lesion-based analysis by Kurkure et al. [7] (4.7 false positives per patient), Shahzad et al. [9] (1.0 to 1.5 false positives per patient), and Wolterink et al. [10] (0.2 false positives per patient). Recently, two studies adopting deep learning-based methods reported high reliability for CAC_auto defined by ICC and kappa values [11,12], which was comparable to that of the present study. However, the investigators did not perform per-lesion analyses. Moreover, the sample sizes of electrocardiogram-synchronized CT scans were relatively small: 87 scans by van Velzen et al. [11] and 511 scans by Martin et al. [12]. Supplementary Table 1 summarizes the previous publications [6-14,24] related to CAC_auto. The present study used the largest electrocardiogram-

synchronized CT dataset (n = 2985) and showed comparable results to those of state-of-the-art studies, as reflected in the ICC (0.99), per-lesion sensitivity (93.3%), and false-positive rate per CT scan (0.11).

Isgum et al. [6] proposed a feature-based method that extracted the size, shape, spatial, and appearance features from calcium candidate objects and applied classifiers to minimize the false-positive rate. Kurkure et al. [7] also utilized features and extended the classifier model in a two-stage mode. These feature-based approaches do not require the spatial information of coronary arteries. However, the results were not comprehensive and required careful selection of several operational parameters for calcium candidate object detection and classification. In another method, Brunner et al. [8] presented an atlas-based approach using coronary artery region models. This method incorporates image transformation to align cardiac volumes across patients and provides coronary artery zones and sections. However, the performance of this atlas-based method was inferior to that of feature-based methods. To combine the benefits of feature- and

atlas-based methods, Shahzad et al. [9] proposed a multi-atlas-based method. Spatial information was obtained from the pairwise registration of the 10 atlas images. However, this multiatlas-based method had a scalability problem. The method was slower when more atlases were included because of the heavy computation of non-rigid registration. In our previous study, when 100 atlases were used, it took approximately 30 minutes [25]. In the present study, to solve the problem of atlas-based methods using registration, we incorporated deep learning–based semantic segmentation to reduce the execution time.

Some researchers have recently proposed deep-learning-based CAC_auto techniques [6,11,12]. Lessmann et al. [13] used a convolutional neural network (CNN) to classify calcium candidate objects. In this approach, two 2.5D CNNs with different receptive fields were consecutively applied. This algorithm can eliminate the feature extraction step in previous approaches and can be trained on large datasets. However, this did not provide spatial information about the coronary artery regions or other surrounding structures. It also requires more computation time to apply voxel-by-voxel classification. Martin et al. [12] presented a multistep deep-learning model. The first step was used to identify and segment the regions, such as the coronary artery, aorta, aortic valve, and mitral valve. The second step classified the voxels as coronary calcium. Zhang et al. [15] reported a multi-step deep learning model, which incorporated finding the calcium candidate region in the axial image and performing a false-positive reduction using shape constraints. The incremental value of this current approach is that it can precisely detect coronary artery regions with a deep learning model based on semantic segmentation in a single step. This method can also provide regional information on the coronary artery and surrounding structures, such as the aorta, ventricular chambers, and myocardium. Therefore, this method can be easily extended to the segmentation of the aortic and mitral valves. On the other hand, the previously used methods share a structural similarity, related to the initial candidate detection and false-positive reduction [12,13,15]. In the first part, a convolutional deep learning model was used to determine whether the region of interest was a coronary artery and it included the aorta, aortic valve, and mitral valve. In the latter part, a convolutional deep learning model was applied to classify whether calcium candidates above 130 HU were coronary calcium. The latter part was added because the number of false positives was large only for the first part.

Although CAC_auto exhibited high reliability in this study, there were still some outliers, as well as false-positive and false-negative errors. The limits of agreement in the high CAC (> 400) group were still high (37.7 ± 197.0) despite a high ICC value. Because the ICC is calculated using variance, the ICC value tends to increase when the data range is wide [26]. Further studies to improve the CAC_auto algorithm are warranted. Two approaches for the improved algorithm were considered. The first approach involves image quality improvement using deep learning. The most frequent reason for mismatched lesions in this study was image noise or motion artifacts. In our previous study [27], we effectively reduced noise in low-dose cardiac CT. The second approach is to improve the training dataset in terms of size and quality. Owing to the nature of AI technology, generalization can be increased when various types of datasets are included during training. In this study, 100 cases of hard-labeled (pixel-by-pixel labeling) data were used for training. We are considering not only adding more training data but also applying data augmentation to capture the variability of the heart or aorta shape due to severe cardiovascular disease. Moreover, a reliable but straightforward verification step for CAC_auto, such as multiple color-coded volume rendering of CAC (Fig. 2), would help shorten the time needed for human confirmation in a CAC_auto workflow. Second, although our study included three different cohort datasets representing three different scenarios and four CT machines, it was still fraught with the limitations of single-center retrospective studies. In addition, a relatively small number (n = 116) of patients with high CAC (> 400) would not be sufficient for AI model validation. Therefore, further multicenter investigations covering larger varieties of CT machines, disease spectrum, and CAC imaging protocols are required. Third, the present study did not evaluate whether CAC_auto streamlined the actual CAC imaging workflow. In follow-up studies, it will be necessary to confirm that CAC_auto can reduce the analysis time while maintaining accuracy when compared with manual work. Finally, since our study used an atlas-based method, theoretically, mitral annular calcification and aortic valve calcification could be automatically evaluated. Considering recent research that reported a link between structural heart disease and non-coronary calcification [28], automatic quantification of valvular or paravalvular calcium would also be promising. Further investigation of the automatic quantification of non-coronary calcifications is warranted.

In conclusion, the atlas-based CAC_auto empowered by

deep learning provided accurate calcium score measurement as compared with manual method and risk category classification, which could potentially streamline CAC imaging workflows.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2021.0148.

Conflicts of Interest

June-Goo Lee owns stocks of Coreline Soft, Co. Ltd., a medical software company in South Korea. The other authors have no relationships to disclose relevant to the content of this paper.

Author Contributions

Conceptualization: Dong Hyun Yang, Young-Hak Kim. Data curation: Dong Hyun Yang, Joon-Won Kang, Hyun Jung Koo. Formal analysis: Dong Hyun Yang, June-Goo Lee, HeeSoo Kim, Heejun Kang. Funding acquisition: Dong Hyun Yang, Young-Hak Kim. Investigation: Dong Hyun Yang, June-Goo Lee, HeeSoo Kim, Heejun Kang. Methodology: Dong Hyun Yang, June-Goo Lee, HeeSoo Kim, Heejun Kang. Project administration: Dong Hyun Yang, Young-Hak Kim. Resources: Dong Hyun Yang, Young-Hak Kim. Software: Dong Hyun Yang, June-Goo Lee, HeeSoo Kim, Heejun Kang. Supervision: Dong Hyun Yang, Young-Hak Kim. Validation: Dong Hyun Yang, Joon-Won Kang, Hyun Jung Koo. Visualization: Dong Hyun Yang, June-Goo Lee, HeeSoo Kim, Heejun Kang. Writing—original draft: Dong Hyun Yang, June-Goo Lee. Writing—review & editing: Joon-Won Kang, Hyun Jung Koo, HeeSoo Kim, Heejun Kang, Young-Hak Kim.

ORCID iDs

June-Goo Lee
    https://orcid.org/0000-0002-1380-6682
HeeSoo Kim
    https://orcid.org/0000-0002-6715-9628
Heejun Kang
    https://orcid.org/0000-0002-0396-2112
Hyun Jung Koo
    https://orcid.org/0000-0001-5640-3835
Joon-Won Kang
    https://orcid.org/0000-0001-6478-0390
Young-Hak Kim
    https://orcid.org/0000-0002-3610-486X

Dong Hyun Yang
    https://orcid.org/0000-0001-5477-558X

REFERENCES

1. 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;15:827-832
2. Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson KE. Coronary calcium score and cardiovascular risk. *J Am Coll Cardiol* 2018;72:434-447
3. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. *J Am Coll Cardiol* 2007;49:378-402
4. Blaha MJ, Mortensen MB, Kianoush S, Tota-Maharaj R, Cainzos-Achirica M. Coronary artery calcium scoring: is it time for a change in methodology? *JACC Cardiovasc Imaging* 2017;10:923-937
5. Hecht HS, Cronin P, Blaha MJ, Budoff MJ, Kazerooni EA, Narula J, et al. 2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: a report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology. *J Cardiovasc Comput Tomogr* 2017;11:74-84
6. Isgum I, Rutten A, Prokop M, van Ginneken B. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. *Med Phys* 2007;34:1450-1461
7. Kurkure U, Chittajallu DR, Brunner G, Le YH, Kakadiaris IA. A supervised classification-based method for coronary calcium detection in non-contrast CT. *Int J Cardiovasc Imaging* 2010;26:817-828
8. Brunner G, Chittajallu DR, Kurkure U, Kakadiaris IA. Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data. *Int J Cardiovasc Imaging* 2010;26:829-838
9. Shahzad R, van Walsum T, Schaap M, Rossi A, Klein S, Weustink AC, et al. Vessel specific coronary artery calcium scoring: an automatic system. *Acad Radiol* 2013;20:1-9
10. Wolterink JM, Leiner T, Takx RA, Viergever MA, Isgum I. Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. *IEEE Trans Med Imaging* 2015;34:1867-1878

11. van Velzen SGM, Lessmann N, Velthuis BK, Bank IEM, van den Bongard DHJG, Leiner T, et al. Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology* 2020;295:66-79

12. Martin SS, van Assen M, Rapaka S, Hudson HT Jr, Fischer AM, Varga-Szemes A, et al. Evaluation of a deep learning–based automated CT coronary artery calcium scoring algorithm. *JACC Cardiovasc Imaging* 2020;13:524-526

13. Lessmann N, van Ginneken B, Zreik M, de Jong PA, de Vos BD, Viergever MA, et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans Med Imaging* 2018;37:615-625

14. Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal* 2016;34:123-136

15. Zhang N, Yang G, Zhang W, Wang W, Zhou Z, Zhang H, 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

16. Kang SJ, Kim YH, Lee JG, Kang DY, Lee PH, Ahn JM, et al. Impact of subtended myocardial mass assessed by coronary computed tomographic angiography-based myocardial segmentation. *Am J Cardiol* 2019;123:757-763

17. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010;29:196-205

18. Shamonin DP, Bron EE, Lelieveldt BP, Smits M, Klein S, Staring M; Alzheimer's Disease Neuroimaging Initiative. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform* 2013;7:50

19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* 2015;arXiv:1409.1556

20. Kim H, Jung J, Kim J, Cho B, Kwak J, Jang JY, et al. Abdominal multi-organ auto-segmentation using 3D-patch-based deep convolutional neural network. *Sci Rep* 2020;10:6204

21. Yang DH, Kang JW, Kim HK, Choe J, Baek S, Kim SH, et al. Association between C-reactive protein and type of coronary arterial plaque in asymptomatic patients: assessment with coronary CT angiography. *Radiology* 2014;272:665-673

22. Yang DH, Kang SJ, Koo HJ, Kweon J, Kang JW, Lim TH, et al. Incremental value of subtended myocardial mass for identifying FFR-verified ischemia using quantitative CT angiography: comparison with quantitative coronary angiography and CT-FFR. *JACC Cardiovasc Imaging* 2019;12:707-717

23. Koo HJ, Kang JW, Oh SY, Kim DH, Song JM, Kang DH, et al. Cardiac computed tomography for the localization of mitral valve prolapse: scallop-by-scallop comparisons with echocardiography and intraoperative findings. *Eur Heart J Cardiovasc Imaging* 2019;20:550-557

24. 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

25. Lee JG, Gumus S, Moon CH, Kwoh CK, Bae KT. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. *Med Phys* 2014;41:092303

26. Mehta S, Bastero-Caballero RF, Sun Y, Zhu R, Murphy DK, Hardas B, et al. Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. *Stat Med* 2018;37:2734-2752

27. Kang E, Koo HJ, Yang DH, Seo JB, Ye JC. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Med Phys* 2019;46:550-562

28. Eleid MF, Foley TA, Said SM, Pislaru SV, Rihal CS. Severe mitral annular calcification: multimodality imaging for therapeutic strategies and interventions. *JACC Cardiovasc Imaging* 2016;9:1318-1337