An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework

Jelmer M. Wolterink[a)]
Image Sciences Institute, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands

Tim Leiner
Department of Radiology, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands

Bob D. de Vos
Image Sciences Institute, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands

Jean-Louis Coatrieux
INSERM, U1099, Rennes F-35000, France; LTSI, Université de Rennes 1, Rennes F-35000, France; and Centre de Recherche en Information Biomédicale Sino-Français (LIA CRIBs), Nanjing 210096, China

B. Michael Kelm
Imaging and Computer Vision, Corporate Technology, Siemens AG, Erlangen 91051, Germany

Satoshi Kondo
Konica Minolta, Inc., Osaka 569-8503, Japan

Rodrigo A. Salgado
Department of Radiology, University Hospital Antwerpen, Edegem 2650, Belgium

Rahil Shahzad
Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands; Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam 3000 CA, The Netherlands; and Quantitative Imaging Group, Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft 2600 GA, The Netherlands

Huazhong Shu
Centre de Recherche en Information Biomédicale Sino-Français (LIA CRIBs), Nanjing 210096, China and Lab of Image Science and Technology, School of Computer Science and Technology, Nanjing 210096, China

Miranda Snoeren
Department of Radiology, Radboud University Medical Center, Nijmegen 6500 HB, The Netherlands

Richard A. P. Takx
Department of Radiology, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands

Lucas J. van Vliet
Quantitative Imaging Group, Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft 2600 GA, The Netherlands

Theo van Walsum
Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam 3000 CA, The Netherlands

Tineke P. Willems
Department of Radiology, University Medical Center Groningen, Groningen 9700 RB, The Netherlands

Guanyu Yang
Lab of Image Science and Technology, School of Computer Science and Technology, Nanjing 210096, China and Centre de Recherche en Information Biomédicale Sino-Français (LIA CRIBs), Nanjing 210096, China

Yefeng Zheng
Imaging and Computer Vision, Corporate Technology, Siemens Corporation, Princeton, New Jersey 08540-6632

Max A. Viergever and Ivana Išgum
Image Sciences Institute, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands

(Received 9 November 2015; revised 18 March 2016; accepted for publication 27 March 2016; published 13 April 2016)

Purpose: The amount of coronary artery calcification (CAC) is a strong and independent predictor of cardiovascular disease (CVD) events. In clinical practice, CAC is manually identified and automatically quantified in cardiac CT using commercially available software. This is a tedious and time-consuming process in large-scale studies. Therefore, a number of automatic methods that require

no interaction and semiautomatic methods that require very limited interaction for the identification of CAC in cardiac CT have been proposed. Thus far, a comparison of their performance has been lacking. The objective of this study was to perform an independent evaluation of (semi)automatic methods for CAC scoring in cardiac CT using a publicly available standardized framework.

Methods: Cardiac CT exams of 72 patients distributed over four CVD risk categories were provided for (semi)automatic CAC scoring. Each exam consisted of a noncontrast-enhanced calcium scoring CT (CSCT) and a corresponding coronary CT angiography (CCTA) scan. The exams were acquired in four different hospitals using state-of-the-art equipment from four major CT scanner vendors. The data were divided into 32 training exams and 40 test exams. A reference standard for CAC in CSCT was defined by consensus of two experts following a clinical protocol. The framework organizers evaluated the performance of (semi)automatic methods on test CSCT scans, per lesion, artery, and patient.

Results: Five (semi)automatic methods were evaluated. Four methods used both CSCT and CCTA to identify CAC, and one method used only CSCT. The evaluated methods correctly detected between 52% and 94% of CAC lesions with positive predictive values between 65% and 96%. Lesions in distal coronary arteries were most commonly missed and aortic calcifications close to the coronary ostia were the most common false positive errors. The majority (between 88% and 98%) of correctly identified CAC lesions were assigned to the correct artery. Linearly weighted Cohen's kappa for patient CVD risk categorization by the evaluated methods ranged from 0.80 to 1.00.

Conclusions: A publicly available standardized framework for the evaluation of (semi)automatic methods for CAC identification in cardiac CT is described. An evaluation of five (semi)automatic methods within this framework shows that automatic per patient CVD risk categorization is feasible. CAC lesions at ambiguous locations such as the coronary ostia remain challenging, but their detection had limited impact on CVD risk determination. © *2016 American Association of Physicists in Medicine.* [http://dx.doi.org/10.1118/1.4945696]

Key words: coronary artery calcification, automatic coronary calcium scoring, cardiac CT, independent method comparison, cardiovascular disease risk, evaluation framework

1. INTRODUCTION

Cardiovascular disease (CVD) is the global leading cause of death.[1] The amount of coronary artery calcification (CAC) is a strong and independent predictor of CVD events.[2–6] Therefore, in clinical practice, CAC is routinely identified and quantified in CT of the heart, most commonly in noncontrast-enhanced ECG-triggered cardiac calcium scoring CT (CSCT). In calcium scoring, a human operator manually identifies CAC lesions in each image slice using commercially available software, which subsequently quantifies the manually identified lesions. Manual identification of CAC is not considered a difficult task, but it is time-consuming and impractical in large-scale (epidemiological) studies. Moreover, the number of cardiac CT exams is expected to rise as recent guidelines recommend CT-based CAC scoring in adults at intermediate and low-to-intermediate risk of a CVD event.[7] As an alternative to manual CAC lesion identification, software that requires no or very limited interaction, i.e., (semi)automatic calcium scoring methods, would reduce the workload of experts and enable large-scale studies.

In the research literature, a number of (semi)automatic methods for coronary calcium scoring in CT have been proposed (Table I). In order to assess the clinical potential of these methods, a detailed comparison of their performance is required. However, a comparison based purely on published results does not provide a reliable assessment for several reasons. First, methods have been evaluated using different data sets, in terms of image acquisition, image reconstruction, and patient inclusion. CAC scoring methods have been developed for diverse CT images of the heart, namely, CSCT,[10,11,16–18,20,21,23] contrast-enhanced coronary CT angiography (CCTA),[9,12,14,19,24,25] a combination of CSCT and CCTA (Ref. 22), or noncontrast-enhanced non-ECG-triggered chest CT.[13,26] These scans may all be used to determine the amount of CAC, but their characteristics pose different challenges. CCTA provides excellent visibility of the coronary arteries, but it may be hard to differentiate between CAC and coronary lumen. On the other hand, while the contrast between CAC and coronary lumen is high in noncontrast-enhanced CT, it is practically impossible to localize the coronary arteries, particularly in non-ECG-triggered chest CT. In addition, images have been acquired on scanners from different vendors, using either electron beam CT (EBCT) or multidetector CT (MDCT), which may yield different image characteristics. The number of included patients per study ranged from 10 to 530, including lung cancer screening participants[26] and patients with known or suspected CVD.[12] Hence, anatomical characteristics and CAC distribution might differ across studies. Second, the CAC scoring task was described and evaluated differently across studies. While most methods identified CAC irrespective of its location in the coronary artery, several studies also performed the more challenging task of CAC labeling per artery.[10,11,16,21] Studies have used slightly different definitions of CAC lesions, including variations in HU threshold and limitations on the required minimum

TABLE I. Published evaluations of methods for automatic coronary calcium scoring, with most recent publications listed first. Reported evaluation: scans required for analysis (scans; noncontrast CSCT, CCTA, chest CT), CT scanner type (CT type; MDCT, EBCT) and vendor (CT vendor; GE, Philips, Siemens, Toshiba), number of test patients (patients), and reported evaluation (evaluation; per lesion, per artery, per patient).

	Scans	CT type	CT vendor	Patients	Evaluation
Schuhbaeck *et al.* (Ref. 8)[a]	CCTA	MDCT	Siemens	44	Patient
Wolterink *et al.* (Ref. 9)	CCTA	MDCT	Philips	50	Patient
Wolterink *et al.* (Ref. 10)	CSCT	MDCT	Philips	530	Lesion, artery, patient
Ding *et al.* (Ref. 11)	CSCT	EBCT, MDCT	GE, Siemens	50	Artery, patient
Ahmed *et al.* (Ref. 12)	CCTA	MDCT	Toshiba	100	Patient
Xie *et al.* (Ref. 13)	Chest CT	—	—	41	Patient
Eilot *et al.* (Ref. 14)	CCTA	MDCT	Philips, Siemens	263	Lesion, patient
Takx *et al.* (Ref. 15)[b]	Chest CT	MDCT	Philips	1749	Patient
Shahzad *et al.* (Ref. 16)	CSCT	MDCT	Siemens	157	Lesion, artery, patient
Wu *et al.* (Ref. 17)	CSCT	—	—	16	Patient
Arnold *et al.* (Ref. 18)	CSCT	MDCT	GE	78	Patient
Mittal *et al.* (Ref. 19)	CCTA	—	—	165	Lesion
Kurkure *et al.* (Ref. 20)	CSCT	EBCT	—	105	Lesion
Brunner *et al.* (Ref. 21)	CSCT	EBCT	GE	30	Lesion, artery
Saur *et al.* (Ref. 22)	CCTA, CSCT	MDCT	Siemens	127	Lesion
Išgum *et al.* (Ref. 23)	CSCT	MDCT	Philips	76	Lesion, patient
Wesarg *et al.* (Ref. 24)	CCTA	MDCT	Siemens	10	Lesion, patient

[a]Method description provided by Dey *et al.* (Ref. 25).
[b]Method description provided by Išgum *et al.* (Ref. 26).

lesion size. Moreover, a range of quantitative evaluation results have been reported, including different correlation coefficients and classification performance measures. Finally, reimplementation of published methods for a comparison using standardized data and evaluation criteria would hardly be feasible due to their complexity.

In this study, we present a publicly available framework for the evaluation of (semi)automatic methods for coronary artery calcium scoring (orCaScore) which was set up to overcome these issues. In this framework, corresponding CSCT and CCTA scans of a stratified set of patients are provided. This set consists of both male and female patients distributed over four CVD risk categories, who were scanned on scanners from four major CT vendors, acquired in four different hospitals. A detailed task description and evaluation criteria were provided. Researchers were invited to analyze the provided exams available within this study using their own research or commercially available (semi)automatic methods. Evaluation was performed by the framework organizers. This work describes the orCaScore framework and an independent comparison of five methods for (semi)automatic CAC scoring.

2. MATERIALS AND METHODS

2.A. orCaScore evaluation framework

The orCaScore evaluation framework is web-based.[27] The website provides a detailed description of the task, available data, and evaluation protocol. Using the website, researchers from academia and industry can register to evaluate their (semi)automatic calcium scoring method. After registration, clinically obtained cardiac CT exams with pairs of CSCT and corresponding CCTA scans may be downloaded. These exams have been divided into a training set and a test set. A reference standard is available to participants for training exams but not for test exams. The task is to identify CAC lesions in the test exams and to optionally label each CAC lesion according to the coronary artery it is located in. For this, participants use their own research or commercial software which is either fully automatic or semiautomatic, i.e., requiring very limited manual interaction. Obtained results are then submitted for evaluation, performed by the study organizers using a standardized set of performance criteria.

The framework was launched at a workshop organized in conjunction with the 2014 Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in Boston, MA. At this workshop, an additional set of test exams was provided for on-site analysis. Potential participants were notified of the framework and workshop by personal invitation, via an image analysis e-mail list and by publicity provided by the MICCAI organizers. This work describes an evaluation of four methods that were presented at the workshop and one method developed by the framework organizers. Here, combined results on both the web-based and the workshop test set are presented. The web-based evaluation framework remains open for future submissions of new or updated methods, results of which will be posted on the website.

2.B. Data

The framework provides cardiac CT exams of patients with both a CSCT and a corresponding CCTA, from four academic hospitals (Antwerp University Hospital, Antwerp, Belgium; Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; University Medical Center Groningen, Groningen, The Netherlands; and University Medical Center

Utrecht, Utrecht, The Netherlands). Patients were considered for inclusion in consecutive scanning order. Those patients with anatomical abnormalities, intracoronary stents, and metal implants as well as CTs showing severe motion artifacts or extremely high levels of noise determined by visual inspection were excluded. Eighteen (nine male, nine female) patients were included per hospital for a total of 72 exams. Patient inclusion was stratified according to CVD risk categories based on Agatston scores, with cutoff points as proposed by Detrano *et al.*[2] (I: 0, II: 1–100, III: 101–300, IV: >300). The web-based training and test exams both included images of 32 patients: one male and one female patient in each CVD risk category from each site. The test exams provided at the workshop included images of eight patients: one patient in category II and one in category IV from each site. This set was limited in size to allow on-site analysis during the workshop.

For each patient, a CCTA and a CSCT scan were acquired in one session as part of clinical routine. Patients were scanned on a multidetector row CT scanner from one of the four different vendors (Lightspeed VCT, software version GMP VCT.26, GE Healthcare, Milwaukee, Wisconsin; Brilliance iCT, software version 3.2.0, Philips Healthcare, Best, The Netherlands; SOMATOM Definition Flash, software version SYNGO CT 2010A, Siemens Healthcare, Forchheim, Germany; and Aquilion ONE, software version 4.93, Toshiba Medical Systems, Otawara, Japan). All acquisitions were synchronized to the diastolic rest period by ECG-triggering, at 70% (GE, Siemens), 75% (Toshiba), or 78% (Philips) of the *R–R* interval. Figure 1 shows example CSCT and CCTA images for four patients with CAC in the left coronary artery tree, who were each scanned on a different scanner.

The CSCTs were acquired with a vendor-specific sequential scanning protocol and a tube voltage of 120 kVp. Images with 0.4×0.4 to 0.5×0.5 mm^2 in-plane resolution and 2.5 mm (GE) or 3 mm (Philips, Siemens, Toshiba) section thickness and increment were reconstructed using vendor-specific software, with the following kernels: standard (GE), XCA (Philips), B35f (Siemens), and FC12 (Toshiba).

The CCTAs were acquired with a vendor-specific sequential scanning protocol, contrast bolus-tracking, and a tube voltage of 100 or 120 kVp. Images with 0.4×0.4 to 0.5×0.5 mm^2 in-plane resolution and 0.625/0.625 mm (GE), 0.9/0.45 mm (Philips), 0.6/0.4 mm (Siemens) and 0.5/0.25 mm (Toshiba) section thickness/increment were reconstructed using vendor-specific software, with the following kernels: standard (GE), XCA (Philips), B30f (Siemens), and FC03 (Toshiba).

Information about the hospitals or scanners from which CTs originated or the distribution of patients with respect to sex or CVD risk category was not provided to study participants.

2.C. Reference standard

Each exam contained a CSCT and a CCTA scan. CSCT is the *de facto* standard for CAC identification and quantification and hence, evaluation was limited to CAC identification in CSCT. Therefore, a reference standard for CAC was set in all provided CSCT scans using annotations by two expert observers: a radiologist (TL) with 12 years of experience in CAC scoring (>5000 CSCT scans) and a research physician (RAPT) with five years of experience in CAC scoring (>1000 CSCT scans). Thresholding at 130 HU identified potential CAC lesions (6-connected components with volume >1.5 mm^3). These included not only CAC but also other high-density lesions such as bony structures and aortic calcifications. Both observers independently identified each CAC lesion and labeled it according to location [left anterior descending (LAD) artery including the left main stem, left circumflex (LCX) artery, or right coronary artery (RCA)]. In a subsequent joint reading session, the observers resolved labeling differences by consensus. Annotations after

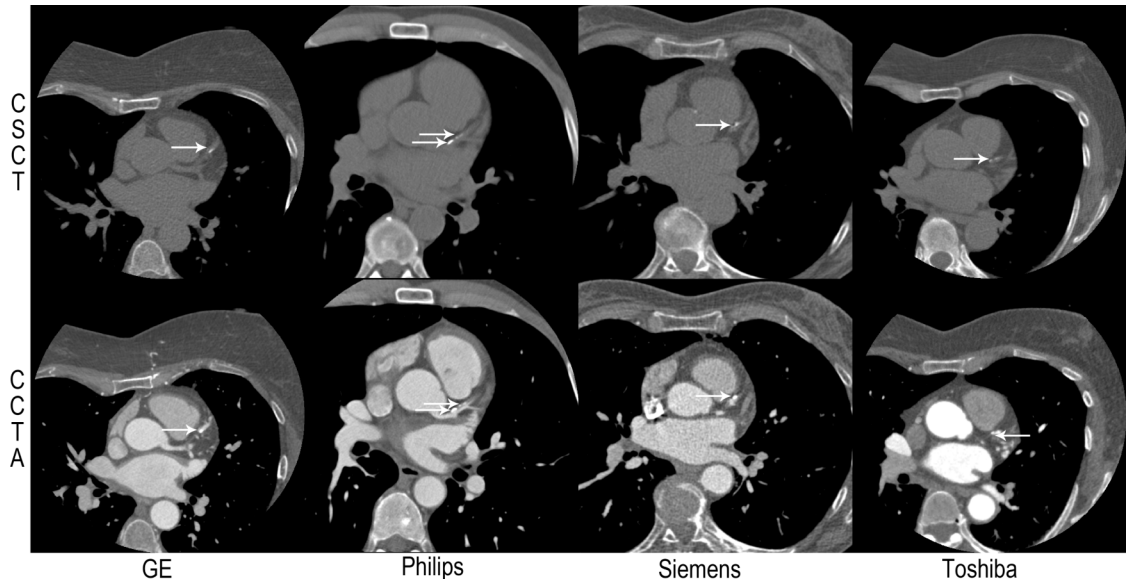

FIG. 1. CSCT and corresponding CCTA scans of four patients contained in the training set, scanned on a GE, Philips, Siemens and Toshiba scanner, respectively. Each of these patients had CAC in the left coronary artery tree, marked with white arrows. Window and level are the same for all images.

consensus served as the reference standard. All annotations were made using custom software (iX Viewer, Image Sciences Institute, Utrecht, The Netherlands) which showed the CSCT scan synchronized with the corresponding CCTA scan for reference.

2.D. Evaluated methods

Four methods (A, B, C, and D) participated in the workshop. A fifth method (E) was developed by the framework organizers, who had access to all study data and the reference standard. Table II lists the required CT scan, user interaction, provided labels for identified lesions and the average processing runtime for each method. All methods required both CSCT and CCTA scans for CAC identification, except for method E, which only used CSCT scans. Methods A, C, and E were automatic, while the other two methods were semiautomatic, requiring manual interaction in all (method D) or a limited number (method B) of exams. The evaluated methods typically identified CAC in two subsequent stages: in the first stage information about the location of the coronary arteries was obtained; in the second stage this information was used to identify CAC lesions. Here, we provide a brief description of each method. A detailed description of each evaluated method is available on the orCaScore website.[27]

2.D.1. Method A—Erasmus Medical Center (Rotterdam, The Netherlands)

CAC was identified by analyzing the test CCTA scan and the test CSCT scan. First, a standardized heart coordinate system and patient-specific probabilistic maps for the occurrence of coronary arteries were obtained for the test CCTA. A probabilistic map contained a probability for each voxel to be part one of the three major coronary arteries (LAD, LCX, and RCA). Eight CCTA atlas scans, each with a heart coordinate system and a probabilistic map, were registered to the test CCTA scan using rigid and subsequent nonrigid registration with *elastix*.[28] The coordinate systems and probabilistic maps were transformed from all atlas CCTA scans to the test CCTA scan and combined through averaging. Next, the test CCTA scan was registered to the test CSCT scan. The heart coordinate system and probabilistic map were transformed from the test CCTA scan to the test CSCT scan. This provided an estimate of the location of the coronary arteries in the test CSCT. Finally, CAC was identified using supervised pattern recognition. All candidate CAC lesions (>130 HU, >1.5 mm^3) in the CSCT were extracted and described by volume, intensity features, coordinates in the heart coordinate system, and the coronary artery probability from the probabilistic map. Candidates were classified as CAC or non-CAC using a nine-nearest neighbor classifier trained with 76 previously acquired CSCT scans. Artery labels of identified CAC lesions were determined using the patient-specific probabilistic map. This method was based on previously described work.[16]

TABLE II. Methods evaluated in this study: scans required for analysis [scans; noncontrast CSCT, CCTA], required user interaction (interaction), labels provided for individual lesions [labels; LAD artery, LCX artery, and RCA] and average runtime per exam (time).

	Scans	Interaction	Labels	Time (min)
Method A	CSCT + CCTA	No interaction required	LAD, LCX, RCA	80
Method B	CSCT + CCTA	Optional correction of automatic ostia detection	LAD, LCX, RCA	8
Method C	CSCT + CCTA	No interaction required	LAD, LCX, RCA	2
Method D	CSCT + CCTA	Manual seed placement in the ascending aorta	LAD, LCX, RCA	3
Method E[a]	CSCT	No interaction required	LAD, LCX, RCA	20

[a]Framework organizer with access to the reference standard.

2.D.2. Method B—Southeast University (Nanjing, China)

CAC was identified by analyzing the test CCTA scan and the test CSCT scan. The test CCTA scan was preprocessed by cylindrical cropping around an initial automatic segmentation of the ascending aorta obtained with circular Hough transforms. Next, the heart and the ascending aorta in the test CCTA were segmented using eight CCTA atlas scans with manual segmentations. Each CCTA atlas was registered to the test CCTA using affine and subsequent nonrigid registration with *elastix*.[28] The heart and aorta segmentations were transformed to the test CCTA scan. Segmentations from all atlas scans were combined through majority voting. Subsequently, the LAD, LCX and RCA centerlines in the test CCTA were tracked using an adapted multiscale vesselness filter.[29,30] To initialize tracking, the coronary ostia were automatically detected and, in case of failed detection, manually annotated. The heart, ascending aorta, and coronary centerline segmentations were transformed from the test CCTA scan to the test CSCT scan in two steps. The test CCTA scan was first registered to the test CSCT scan using affine and subsequent nonrigid registration, and the heart segmentation was mapped from the test CCTA onto the test CSCT scan. Then, an additional nonrigid registration was applied, using the transformed heart segmentation in the test CSCT as a mask. Subsequently, the ascending aorta segmentation and coronary artery centerlines were also propagated from the CCTA to the CSCT. A morphological dilation was applied to the coronary centerlines in the test CSCT and the intersection of this dilation and the heart segmentation was considered a coronary artery mask. The aortic segmentation was subtracted from this mask to exclude potential aortic calcifications. CAC was identified using supervised pattern recognition. Only candidate CAC lesions (>130 HU, <600 voxels) contained in the coronary artery mask of the CSCT were extracted and described by volume, intensity features, and location in the heart coordinate system. Candidates were classified as CAC or non-CAC by means of a support vector machine trained with the training set provided in the framework. Identified CAC was assigned

to an artery using the previously obtained coronary artery labels.

2.D.3. Method C—Siemens Corporate Technology (Princeton, NJ)

CAC was identified by analyzing the test CCTA scan and the test CSCT scan. First, in both the CCTA scan and the CSCT scan, the pericardium and aortic root were automatically segmented. In both scans, marginal space learning was used to estimate the position, orientation, and size of the heart.[31] Mean shapes for the pericardium and aortic root, based on a set of example shapes, were aligned with the estimated pose as initial segmentations. These were then refined using an active shape model. The sternum and ribs were explicitly segmented and subtracted from the obtained pericardium segmentation. LAD, LCX, and RCA coronary artery centerlines in the test CCTA scan were automatically extracted by a combination of model-driven and data-driven methods.[32] These centerlines were propagated from the test CCTA onto the test CSCT, by aligning the coronary root and pericardium contours in both scans using pointwise correspondence. The transformation field within the pericardium was interpolated using a thin-plate-spline model. A coronary artery mask in the CSCT was determined as the intersection of the pericardium segmentation and the coronary artery centerlines, dilated by 1.5 mm. The aortic root was excluded from this mask. Finally, CAC was identified using supervised pattern recognition. Only candidate CAC lesions (>130 HU, >3 voxels, 6-connected) that were at least partially in the coronary artery mask of the CSCT were extracted and described with volume, intensity features, distance to the closest centerline, and coordinates in a heart coordinate system determined by the pericardium segmentation. Candidates were classified as CAC or non-CAC by means of a random forest classifier trained with the training set provided in the framework. Identified CAC was assigned to an artery using the previously obtained coronary artery labels.

2.D.4. Method D—Konica Minolta, Inc. (Osaka, Japan)

CAC was identified by analyzing the test CCTA scan and the test CSCT scan. First, a Gaussian filter was applied to the axial slices in the test CCTA. The lung area in this scan was identified based on HU values and removed. Next, coronary arteries in the test CCTA scan were segmented using region growing. To initialize this process, a point in the ascending aorta was manually identified. A HU threshold was determined based on the intensity of voxels surrounding this point. Subsequent region growing using this threshold identified connected aortic and coronary voxels. The top-most slice in which the identified component appeared disconnected determined the location of the left coronary trunk. The right coronary trunk was identified using an analogous strategy. Subsequently, a multiscale vesselness filter was used to enhance vessels in the test CCTA image.[29] The method then extracted the left and right coronary trees separately by region growing based on intensities and the vesselness values. Next, the test CCTA scan was registered to the test CSCT scan using nonrigid registration. The extracted coronary arteries were transformed from the test CCTA onto the test CSCT to create a coronary artery mask in the test CSCT. A CAC candidate mask was generated by thresholding the test CSCT at 130 HU. Finally, voxels in the intersection of the coronary artery mask and the candidate mask were identified as CAC.

2.D.5. Method E—University Medical Center Utrecht (Utrecht, The Netherlands)

CAC was identified by analyzing only the test CSCT scan. First, the position of the LAD, LCX, and RCA coronary artery centerlines in the test CSCT scan was estimated. For this, ten CSCT atlas scans with LAD, LCX, and RCA centerline annotations were used. The centerlines had previously been manually annotated in corresponding atlas CCTA scans and transformed to the atlas CSCT scans. Each CSCT atlas scan was registered to the test CSCT scan using affine and subsequent elastic registration with *elastix*.[28] The obtained transformation was used to map the centerlines from the atlas CSCT onto the test CSCT. Next, for LAD, LCX, and RCA separately, a centerline was estimated by iterative fusing of the propagated centerlines. In each iteration, the artery with the largest Fréchet distance to the geometric median was discarded. The final coronary artery estimate was the geometric median of the remaining centerlines after termination. Finally, CAC was identified using supervised pattern recognition. All candidate CAC lesions (>130 HU, 6-connected, >1.5 mm^3) in the CSCT were extracted and described by volume, shape features, intensity features, and location features based on the estimated coronary artery centerlines. Candidate lesions were classified as CAC in the LAD, LCX, or RCA or as non-CAC by a multiclass randomized decision forest trained with 237 previously acquired CSCT scans. This method was developed by the framework organizers, who had access to the reference annotations of all data sets. A detailed description of this method is provided in Ref. 10.

2.E. Evaluation

CAC identified by the (semi)automatic methods was compared with the reference standard. In addition, to evaluate the performance of human experts, CAC identified by the two expert observers prior to the consensus reading was compared with the reference standard.

Lesions were defined as 6-connected components of voxels >130 HU with volume >1.5 mm^3. The volume of a lesion (in mm^3) was determined as the number of voxels in the lesion multiplied by the voxel volume. The detection of CAC lesions was evaluated using sensitivity, positive predictive value (PPV) and the F_1 score, i.e., the harmonic average of sensitivity and PPV [$2\cdot(\text{sensitivity}\cdot\text{PPV})/(\text{sensitivity}+\text{PPV})$], weighted by both the total number and total volume of lesions in the test set. For a statistical pairwise comparison of the performance of methods and observers on the detection of CAC lesions, McNemar's test was performed, with Bonferroni correction for 21 pairwise comparisons so that $\alpha = 0.05/21$

= 0.0024.[33] The determination of per patient CAC volume was evaluated using the intraclass correlation coefficient (ICC) for absolute agreement with the reference standard.

For each lesion, in addition to CAC volume, the Agatston score was determined as

$$\sum_{s \in S} a_s \cdot d_s, \tag{1}$$

where S is the set of image slices containing the lesion, a_s is the total lesion area (in mm^2) in s, and d_s is a density factor determined by the lesion's maximum attenuation in s (130–199 HU: 1, 200–299 HU: 2, 300–399 HU: 3, ≥400 HU: 4).[34] Because the Agatston score is based on images with 3 mm section thickness and increment, a linear correction factor 2.5/3.0 was applied for GE images with section thickness and increment of 2.5 mm. Per patient Agatston scores were determined by addition of lesion Agatston scores. To compare per patient Agatston scores with the reference standard, Bland–Altman plots were used. In addition, a CVD risk category was computed per patient based on the Agatston scores (I: 0, II: 1–100, III: 101–300, IV: >300) and agreement with the reference standard was computed using linearly weighted Cohen's kappa. Finally, to evaluate per artery CAC volume determination, the ICC for absolute agreement with the reference standard was computed.

The evaluation was implemented in MATLAB (Version R2013a, Mathworks, Natick, MA, USA).

3. RESULTS

3.A. Detection of CAC lesions

The 40 test exams contained a total number of 316 CAC lesions, corresponding to 9018.5 mm^3. Sensitivity for individual lesion identification ranged from 52% to 94%, corresponding to a sensitivity of 57%–99% for CAC volume. The PPV ranged from 65% to 96% for lesion identification and from 69% to 96% for CAC volume identification. The F_1 score ranged from 57% to 95% for lesion identification and from 62% to 97% for CAC volume identification. Detailed results obtained by all methods and by the two expert observers are listed in Table III. For all methods as well as for the experts, the sensitivity was higher for CAC volume than for CAC lesions. This indicates that correctly identified CAC lesions were larger in volume than missed lesions.

Figure 2 shows F_1 scores for lesion identification obtained by the evaluated methods and the two expert observers. All pairwise differences were analyzed using McNemar's test with Bonferroni correction for multiple comparisons. The null hypothesis that a pair of methods or observers performed equally well held for three pairs: methods C and E (p = 1.0000), method B and observer 1 (p = 0.0315), and the two observers (p = 0.0094). All other pairs yielded $p < 0.05/21$; one method or observer in the pair significantly outperformed the other.

Figure 3 shows confusion matrices comparing reference and (semi)automatically determined labels for all lesions in the test exams. All methods made more false negative (FN) errors than false positive (FP) errors (on average 0.5–3.8 FN errors vs 0.3–2.2 FP per patient). However, for all methods, FP errors were typically larger than FN errors (on average 19.2–60.8 mm^3 per FP error vs 5.2–10.0 mm^3 per FN error). Figure 4 shows typical FN and FP errors. Eight FN errors were common to all methods. With the exception of one larger lesion in the distal LCX [Fig. 4(a)], these lesions were small in volume, mostly located in the proximal LAD, distal LCX, and distal RCA [Fig. 4(b)]. No single FP error was common to all methods, but six identical FP errors were made by three methods. Five of these were located at the right coronary ostium [Fig. 4(c)] and one at the left coronary ostium [Fig. 4(d)].

TABLE III. Sensitivity, positive predictive value (PPV), and F_1 score for the detection of CAC lesions, weighted by number of lesions (lesions) and by volume in mm^3 (volume), achieved by each method and the two expert observers. The highest value in each column is shown in boldface, separately for (semi)automatic methods and observers.

	Sensitivity (%)		PPV (%)		F_1 score (%)	
	Lesions	Volume	Lesions	Volume	Lesions	Volume
Method A[a]	62.3	85.3	87.9	93.7	73.9	89.3
Method B[b]	**94.0**	**98.9**	**95.5**	94.8	**94.7**	**96.8**
Method C[a]	83.9	94.9	95.3	93.8	89.2	94.3
Method D[b]	51.6	57.3	64.7	68.5	57.4	62.4
Method E[a,c]	84.5	93.5	95.0	**95.9**	89.5	94.7
Observer 1	94.3	98.6	**99.3**	**98.6**	96.8	**98.6**
Observer 2	**99.1**	**99.9**	99.1	95.5	**99.1**	97.5

[a]Automatic method.
[b]Semiautomatic method.
[c]Framework organizer with access to the reference standard.

3.B. CAC scoring per patient

Median (interquartile range) reference CAC volume per patient was 84.4 (6.6–327.4) mm^3. Table IV lists the two-way ICC for absolute agreement with the reference standard for both the (semi)automatic methods and the two expert observers. The ICC ranged from 0.60 (0.36–0.76) to 0.99 (0.99–1.00) for the evaluated methods. For the two observers, the ICC values were 1.00 (1.00–1.00) and 0.98 (0.97–1.00), respectively.

Median (interquartile range) reference Agatston score per patient was 72.0 (3.9–361.5). The Bland–Altman plots in

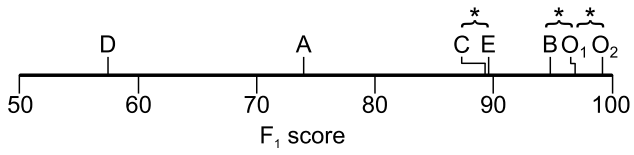

FIG. 2. F_1 scores for the identification of CAC lesions in the test set, for evaluated methods A–E and the two observers O_1 and O_2. The F_1 score was computed as the harmonic average of sensitivity and PPV [2 · (sensitivity · PPV)], weighted by number of lesions. Pairs indicated by a brace and an asterisk were found not to be statistically different by McNemar's test with Bonferroni correction for 21 comparisons ($\alpha = 0.05/21$). All other pairs performed significantly differently.

Automatic

Reference	FN	LAD	LCX	RCA	Total
FP	-	5	7	15	27
LAD	38	**90**	1	0	129
LCX	27	2	**40**	1	70
RCA	54	0	0	**63**	117
Total	119	97	48	79	

a) Method A

Semi-automatic

Reference	FN	LAD	LCX	RCA	Total
FP	-	3	0	11	14
LAD	7	**119**	3	0	129
LCX	7	1	**57**	5	70
RCA	5	0	0	**112**	117
Total	19	123	60	128	

b) Method B

Automatic

Reference	FN	LAD	LCX	RCA	Total
FP	-	6	0	7	13
LAD	11	**108**	10	0	129
LCX	19	7	**42**	2	70
RCA	21	0	0	**96**	117
Total	51	121	52	105	

c) Method C

Semi-automatic

Reference	FN	LAD	LCX	RCA	Total
FP	-	5	1	83	89
LAD	51	**70**	8	0	129
LCX	32	3	**30**	5	70
RCA	70	0	0	**47**	117
Total	153	78	39	135	

d) Method D

Automatic

Reference	FN	LAD	LCX	RCA	Total
FP	-	6	1	7	14
LAD	10	**109**	10	0	129
LCX	16	2	**51**	1	70
RCA	23	0	0	**94**	117
Total	49	117	62	102	

e) Method E

Manual

Reference	FN	LAD	LCX	RCA	Total
FP	-	1	0	1	2
LAD	5	**122**	2	0	129
LCX	5	1	**64**	0	70
RCA	8	0	7	**102**	117
Total	18	124	71	103	

f) Observer 1

Manual

Reference	FN	LAD	LCX	RCA	Total
FP	-	0	0	3	3
LAD	1	**126**	2	0	129
LCX	1	0	**67**	2	70
RCA	1	0	1	**115**	117
Total	3	126	70	120	

g) Observer 2

FIG. 3. Confusion matrices for lesion labeling per coronary artery. Each cell shows the number of lesions assigned a coronary artery label in the reference standard (rows) and by the (semi)automatic methods or observers (columns). The diagonal cells show the number of lesions for which the labeling by the reference standard and (semi)automatic methods or observers agreed. The off-diagonal cells show the number of lesions where they disagreed. Labels were LAD artery, LCX artery, RCA, and negative (Neg). Note that the cells in the first rows and columns show false positive and false negative errors, respectively.

Fig. 5 show differences between the reference standard and the (semi)automatic methods, as well as between the reference standard and the two expert observers. Two patients with calcifications close to the coronary ostia and one patient with CAC lesions in the distal parts of the coronary arteries are marked in each plot.

Figure 6 shows confusion matrices for CVD risk based on Agatston scores. Linearly weighted Cohen's kappa ranged from 0.80 to 1.00. The patients assigned to an incorrect risk category were, with one exception, assigned to a neighboring category. Overall, 26/40 (65%) patients were assigned to their correct risk category by all methods.

3.C. CAC scoring per artery

The test exams contained 129/316 (41%) LAD, 70/316 (22%) LCX, and 117/316 (37%) RCA CAC lesions. The results in Fig. 3 show that the methods assigned 88%–98% of correctly identified CAC lesions to their reference artery, while this number was 97% and 98% for the observers. Up

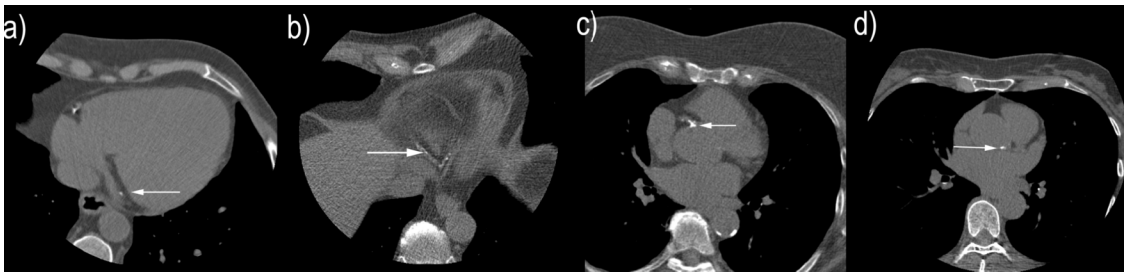

FIG. 4. Examples of typical false positive and false negative errors made by the evaluated methods. (a) A CAC lesion in the distal left circumflex artery which was missed by all methods. (b) Two CAC lesions in the distal right coronary artery which were missed by all methods. (c) A large aortic calcification at the right coronary ostium which was incorrectly detected as CAC by three methods. (d) An aortic calcification at the left coronary ostium which was incorrectly detected as CAC by three methods.

TABLE IV. The two-way ICC for absolute agreement with 95% confidence interval between reference and obtained CAC volume per patient (patient) and per coronary artery [LAD artery, LCX artery, and RCA] shown for each method and for the two expert observers. The highest value in each column is shown in boldface, separately for (semi)automatic methods and observers.

	Patient	LAD	LCX	RCA
Method A[a]	0.97 (0.94–0.98)	0.98 (0.96–0.99)	0.95 (0.91–0.98)	0.90 (0.81–0.94)
Method B[b]	**0.99 (0.99–1.00)**	**1.00 (1.00–1.00)**	**1.00 (0.99–1.00)**	**0.96 (0.93–0.98)**
Method C[a]	0.98 (0.97–0.99)	0.94 (0.88–0.97)	0.87 (0.77–0.93)	0.96 (0.92–0.98)
Method D[b]	0.60 (0.36–0.76)	0.33 (0.04–0.57)	0.57 (0.32–0.74)	0.51 (0.25–0.71)
Method E[a,c]	0.99 (0.97–0.99)	0.98 (0.97–0.99)	0.98 (0.96–0.99)	**0.96 (0.93–0.98)**
Observer 1	**1.00 (1.00–1.00)**	**1.00 (1.00–1.00)**	0.99 (0.98–0.99)	**0.98 (0.97–0.99)**
Observer 2	0.98 (0.97–1.00)	**1.00 (1.00–1.00)**	**1.00 (1.00–1.00)**	0.90 (0.83–0.95)

[a]Automatic method.
[b]Semiautomatic method.
[c]Framework organizer with access to the reference standard.

to 25% of all errors were caused by assigning CAC lesions to the incorrect artery. Figure 7 shows examples of CAC lesions that were assigned to the incorrect coronary artery by multiple methods. CAC lesions in the LAD were incorrectly assigned to the LCX [Fig. 7(a)], while CAC lesions in the LCX were incorrectly assigned to the LAD [Fig. 7(b)]. One lesion was labeled as LCX in the reference standard, but identified as RCA by three methods [Fig. 7(c)]. Retrospective expert inspection of the reference standard revealed that this lesion was located in the RCA. No labeling errors were made between the LAD and the RCA.

Median (interquartile range) reference CAC volume per artery was 52.0 (7.1–189.2), 1.7 (0.0–105.3), and 6.1 (0.0–69.2) mm^3 for LAD, LCX, and RCA, respectively. Table IV lists the two-way ICC for absolute agreement with the reference standard for the (semi)automatic methods and the two expert observers. Three methods obtained a higher agreement for RCA volume than Observer 2, who incorrectly assigned a large aortic calcification at the right coronary ostium to the RCA.

3.D. Performance on scans acquired with different CT scanners

Test exams from the GE, Philips, Siemens, and Toshiba scanner contained 102, 72, 86, and 56 CAC lesions with total CAC volumes 2772.8, 3063.0, 1613.2, and 1569.4 mm^3, respectively. Figure 8 shows the F_1 score for lesion identification separately for scans acquired on each scanner, as well as for the total test set, for the evaluated methods A–E and the two expert observers. This provides an indication of the performance of the evaluated methods across scans from different vendors and shows the effect of each scanner on overall performance of the methods. Note that the number of scans included per scanner was small, and it is likely that other factors than only the scanner vendor affected the differences in performance. Hence, this comparison does not provide sufficient statistical power to generalize the results to other datasets or clinical scenarios.

4. DISCUSSION

We have presented an independent evaluation framework for (semi)automatic methods performing coronary calcium scoring in cardiac CT. The framework provides clinical exams consisting of a CSCT scan with a corresponding CCTA scan and standardized criteria for performance evaluation. Five methods were evaluated within this framework. All methods achieved linearly weighted Cohen's kappa ≥0.80 with the reference standard for patient CVD risk categorization. On a lesion level, some methods can identify CAC in CSCT with a sensitivity and positive predictive value close to that of expert observers. Nevertheless, all evaluated methods made common errors, typically at the coronary ostia and in distal segments of the coronary arteries.

Three of the evaluated methods were fully automatic, requiring no user interaction. The other two methods were semiautomatic, requiring limited user interaction to manually initialize or correct automatic coronary artery extraction. Although this level of expert interaction is substantially lower than in manual CAC scoring as performed in clinical settings, it might reduce the potential of these methods for large studies or screening settings.[35] The focus of the evaluation criteria was on the accuracy of CAC identification and not on time required for the analysis; hence methods were generally not optimized for speed. Consequently, the automatic methods took up to 80 min, which exceeds the time manual scoring would take. However, these methods required no interaction and CAC scores may thus be obtained "off-line" without additional workload for the clinician.

Because CSCT is the standard for CAC scoring in the clinic, methods were evaluated on their ability to identify CAC in CSCT. However, for each CSCT, a corresponding CCTA scan was provided. Four methods (A–D) used this CCTA scan in addition to the CSCT. In clinical practice, determining a CAC score based on only CSCT could be advantageous, as CCTA is not always acquired in conjunction with CSCT.[36] However, information in the CCTA could improve the identification of CAC, due to increased visibility of the coronary arteries in CCTA compared to CSCT. Hence, ideally, a method would be able to process stand-alone CSCTs, as well as pairs of CSCT and CCTA scans when both are available. Method A used the test CCTA as an intermediate for registration of the test CSCT and their atlas CCTA and could be adapted to use only test CSCT, as demonstrated in previous work.[16] Conversely, method E, which now used only CSCT, could be extended to also analyze CCTA. Methods B–D relied on coronary artery segmentation in the patient CCTA, which might limit their performance when no CCTA is available.

Clinical CAC scoring commonly provides a single measure for patient CAC burden. However, it has been shown that

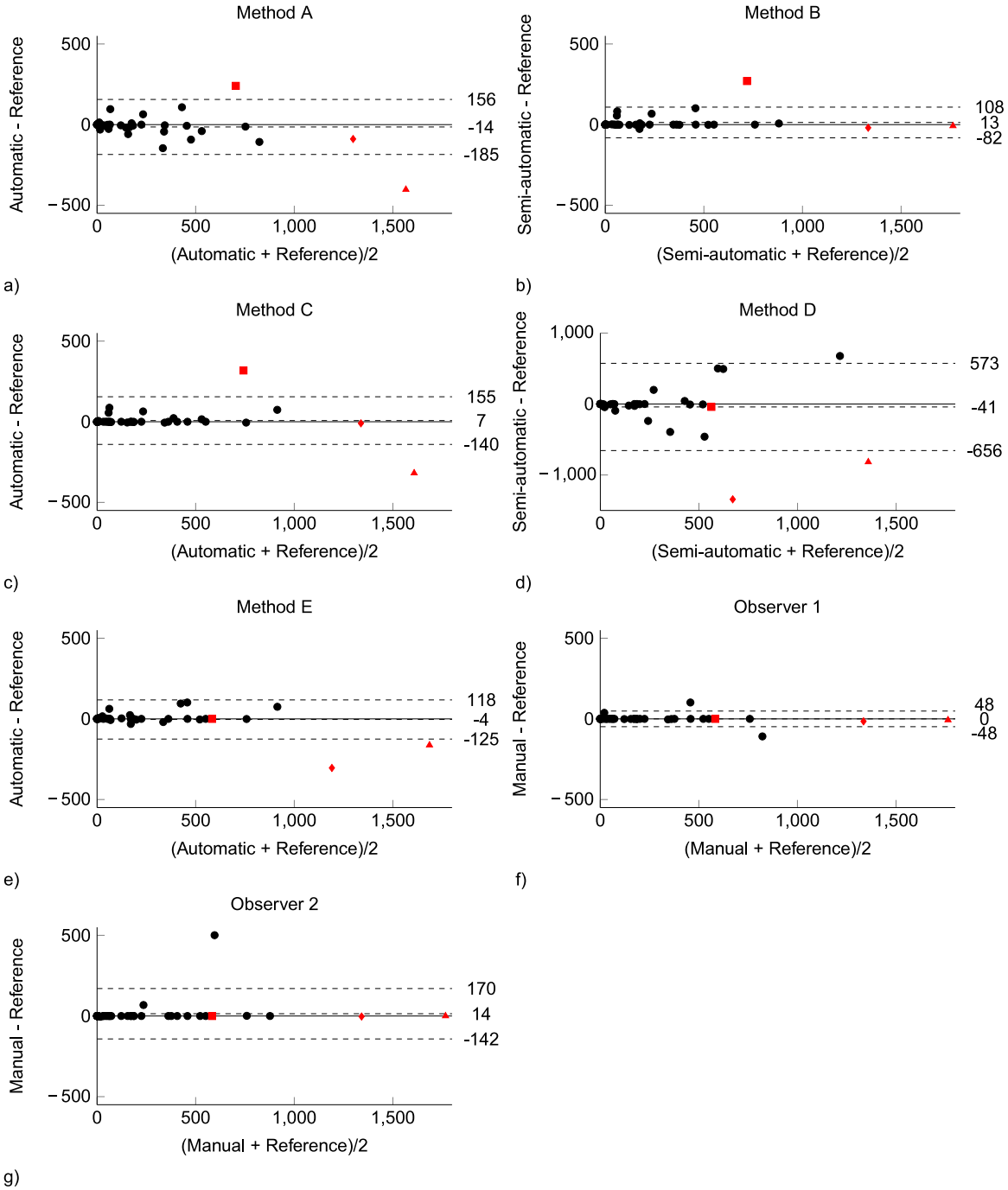

FIG. 5. Bland–Altman plots comparing per patient reference Agatston scores with (semi)automatic methods and expert observers. Three patients for whom the reference standard and the evaluated methods showed the largest disagreement are marked in red in all plots: one patient with a large CAC lesion at the boundary of the ascending aorta and the right coronary artery (RCA, square), one patient with large CAC lesions in the acute marginal branch of the RCA (triangle), and one patient with extensive CAC in the distal left circumflex artery and the RCA (diamond). Dashed lines represent the mean and the limits of agreement of differences (±1.96 SD).

the distribution of CAC over arteries is associated with total coronary plaque burden[37] and is highly predictive of CVD events[38] or future coronary revascularization.[39] Hence, each method's ability to determine this distribution was evaluated. The methods assigned the vast majority (88%–98%) of correctly identified CAC lesions to the correct artery. The evaluated methods identified CAC lesions in the LAD most accurately, followed by those in the LCX and RCA. Reduced performance in the RCA might be explained by the presence of CAC in distal segments, which is more common in the RCA than in the LAD and LCX.[38] The identification of distal CAC in the RCA might be challenging for artery-tracking methods such as B, C, and D, which commonly extract proximal parts of the coronary artery better than distal parts,[40] as well as for multiatlas based methods such as A and E, which are sensitive to anatomical variations in the distal coronary arteries. In addition, differences between identification of CAC in the LAD and in the LCX or RCA might be explained by the final

Automatic

Reference	I	II	III	IV	Total
I	**5**	3	0	0	8
II	1	**10**	1	0	12
III	1	0	**8**	0	8
IV	0	0	1	**11**	12
Total	6	13	10	11	40

a) Method A

Semi-automatic

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	0	**11**	1	0	12
III	0	0	**8**	0	8
IV	0	0	0	**12**	12
Total	8	11	9	12	40

b) Method B

Automatic

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	1	**10**	1	0	12
III	0	0	**8**	0	8
IV	0	0	0	**12**	12
Total	9	10	9	12	40

c) Method C

Semi-automatic

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	2	**10**	0	0	12
III	0	1	**6**	1	8
IV	1	0	3	**8**	12
Total	11	11	9	9	40

d) Method D

Automatic

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	0	**12**	0	0	12
III	0	0	**8**	0	8
IV	0	0	0	**12**	12
Total	8	12	8	12	40

e) Method E

Manual

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	0	**12**	0	0	12
III	0	0	**8**	0	8
IV	0	0	0	**12**	12
Total	8	12	8	12	40

f) Observer 1

Manual

Reference	I	II	III	IV	Total
I	**8**	0	0	0	8
II	0	**12**	0	0	12
III	0	0	**8**	0	8
IV	0	0	0	**12**	12
Total	8	12	8	12	40

g) Observer 2

FIG. 6. Confusion matrices for CVD risk categorization based on the Agatston score. The diagonal cells show correctly categorized patients. The cells above and below the diagonal show patients whose CVD risk was overestimated or underestimated, respectively. The following linearly weighted Cohen's kappa scores were obtained: (a) 0.88, (b) 0.98, (c) 0.96, (d) 0.80, (e) 1.00, (f) 1.00, and (g) 1.00.

supervised classification step common to methods A, B, C, and E, which requires a sufficient number of training samples. The 32 provided training CTs contained more LAD than LCX or RCA CAC lesions, which potentially limits each method's ability to distinguish CAC lesions in the LCX and RCA.

The provided CTs originated from four different CT scanners. Even though scanning protocols for CSCT are standardized,[41] CTs acquired with different scanners have been shown to yield significantly different CAC scores.[42] The results showed that methods A, B, C, and E obtained similar F_1 scores for CTs originating from different scanners. Methods A, B, and E were developed with CTs that were not provided in the framework, acquired with Philips (method E) or Siemens (methods A and B) CT scanners. This could potentially lead to a positive bias toward CTs from the same scanner vendor. However, the results showed only a slightly positive bias toward Philips for method E and even a slightly negative bias toward Siemens for methods A and B. Hence, the results did not indicate a bias toward the scanners used in method development. Method D failed to identify CAC in two scans acquired with the GE scanner and made several large FP errors on another scan acquired with the GE scanner. Hence, its performance on scans obtained with the GE scanner was substantially lower than on scans obtained with the other

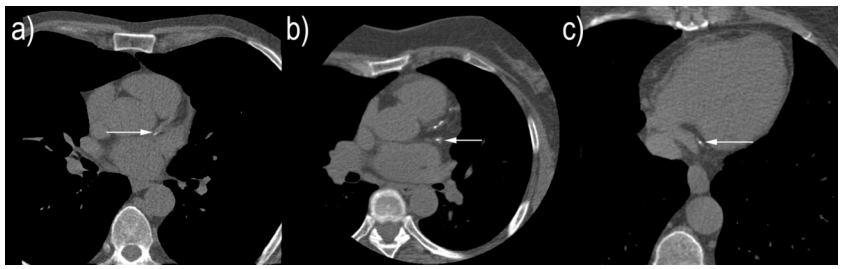

FIG. 7. Examples of typical labeling errors made by the evaluated methods. (a) A CAC lesion in the LAD artery which was assigned to the LCX artery by three methods. (b) A CAC lesion in the LCX which was assigned to the LAD by three methods. (c) A CAC lesion which was labeled as LCX in the reference standard, but assigned to the right coronary artery (RCA) by three methods. Retrospective analysis by an expert observer revealed that this lesion was indeed located in the RCA.

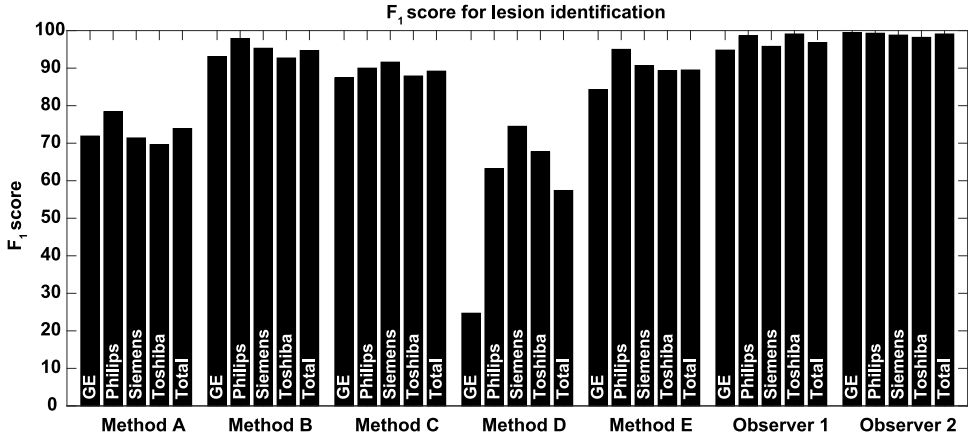

FIG. 8. F_1 score for lesion identification on scans obtained on four different scanners (GE, Philips, Siemens, Toshiba), and the overall test set (Total), for the evaluated methods A–E, and the two expert observers. The F_1 score was computed as the harmonic average of sensitivity and PPV [$2 \cdot (\text{sensitivity} \cdot \text{PPV})$], weighted by number of lesions.

scanners. This occurred due to failure to detect the left and right coronary ostia, and hence failure to extract the coronary artery tree. The extent of differences between scanners, the limited number of test exams, and other factors such as patient characteristics and patient setup during scanning prohibits us to make conclusions. In future work, the provided CTs could be supplemented with a large number of scans originating from a range of different scanners and scanner vendors allowing an in-depth analysis of the performance of evaluated methods on diverse data.

Large-scale clinical trials have shown that asymptomatic patients with zero CAC have an excellent CVD prognosis, and that the identification of these patients might prevent unnecessary healthcare costs.[43] The evaluated methods were able to differentiate between zero CAC and nonzero CAC patients. Four out of five methods correctly identified all eight patients without CAC, while the fifth method overestimated CAC in three of these patients owing to small FP errors. Conversely, two methods correctly identified all 32 patients with positive CAC, while the remaining three methods incorrectly assigned a zero CAC score to at most three patients. These results indicate that (semi)automatic CAC scoring might be used to identify zero CAC patients.

Although CAC is typically identified and quantified in dedicated cardiac CT scans, it may also be scored in other CT scans visualizing the heart. This is especially interesting in non-ECG-triggered low-dose chest CT acquired in lung cancer screening. However, automatic CAC scoring in chest CT images poses different challenges, because these scans may contain severe motion artifacts and high levels of noise. Hence, dedicated methods have been developed for this purpose[13,26] and it is not likely that the methods evaluated in this study would be directly applicable to chest CT. The methods would need to be adjusted to low-dose chest CT, e.g., the requirement of additional CCTA scans for methods A, B, C, and D should be removed.

Scans included in this study provide a set of representative cardiac CT scans. Patients were balanced over gender, CVD risk, and scanning site. Scans with anatomical abnormalities, intracoronary stents, metal implants severe motion artifacts, or extremely high levels of noise determined by visual inspection were excluded. Even though calcium scores determined in such images may not be reliable or scoring might not be possible, in future work the provided CTs could be supplemented with exams containing such abnormalities, to evaluate whether methods would be able to recognize such scans.

5. CONCLUSION

A publicly available standardized framework for the evaluation of (semi)automatic methods for CAC identification in cardiac CT is described. An evaluation of five (semi)automatic methods within this framework shows that automatic per patient CVD risk categorization is feasible. CAC lesions at ambiguous locations such as the coronary ostia remain challenging, but their detection had limited impact on CVD risk determination.

ACKNOWLEDGMENT

This study has been financially supported by PIE Medical Imaging B.V.

[a]Electronic mail: j.m.wolterink@umcutrecht.nl

[1]World Health Organization, WHO Fact Sheet N17: Cardiovascular Diseases, 2013.
[2]R. Detrano *et al.*, "Coronary calcium as a predictor of coronary events in four racial or ethnic groups," N. Engl. J. Med. **358**, 1336–1345 (2008).
[3]J. Yeboah, R. L. McClelland, T. S. Polonsky, G. L. Burke, C. T. Sibley, D. O'Leary, J. J. Carr, D. C. Goff, P. Greenland, and D. M. Herrington, "Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals," JAMA **308**, 788–795 (2012).
[4]A. Becker, A. Leber, C. Becker, and A. Knez, "Predictive value of coronary calcifications for future cardiac events in asymptomatic individuals," A. Heart J. **155**, 154–160 (2008).
[5]J. A. Rumberger, B. H. Brundage, D. J. Rader, and G. Kondos, "Electron beam computed tomographic coronary calcium scanning: A review and guidelines for use in asymptomatic persons," Mayo Clin. Proc. **74**, 243–252 (1999).
[6]K. Nasir and M. Clouse, "Role of nonenhanced multidetector CT coronary artery calcium testing in asymptomatic and symptomatic individuals," Radiology **264**, 637–649 (2012).

[7]D. C. Goff, D. M. Lloyd-Jones, G. Bennett, S. Coady, R. B. D'Agostino, R. Gibbons, P. Greenland, D. T. Lackland, D. Levy, C. J. O'Donnell, J. G. Robinson, J. S. Schwartz, S. T. Shero, S. C. Smith, P. Sorlie, N. J. Stone, and P. W. F. Wilson, "2013 ACC/AHA guideline on the assessment ofcardiovascular risk: A report of the American College of Cardiology/American Heart Association task force on practice guidelines," J. Am. Coll. Cardiol. **63**, 2935–2959 (2014).

[8]A. Schuhbaeck, Y. Otaki, S. Achenbach, C. Schneider, P. Slomka, D. S Berman, and D. Dey, "Coronary calcium scoring from contrast coronary CT angiography using a semiautomated standardized method," J. Cardiovasc. Comput. Tomogr. **9**, 446–453 (2015).

[9]J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Automatic coronary calcium scoring in cardiac ct angiography using convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Lecture Notes in Computer Science, edited by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Springer International, Switzerland, 2015), pp. 589–596.

[10]J. M. Wolterink, T. Leiner, R. A. P. Takx, M. A. Viergever, and I. Isgum, "Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection," IEEE Trans. Med. Imaging **34**, 1867–1878 (2015).

[11]X. Ding, P. J. Slomka, M. Diaz-Zamudio, G. Germano, D. S. Berman, D. Terzopoulos, and D. Dey, "Automated coronary artery calcium scoring from non-contrast CT using a patient-specific algorithm," Proc. SPIE **9413**, 94132U (2015).

[12]W. Ahmed, M. A. de Graaf, A. Broersen, P. H. Kitslaar, E. Oost, J. Dijkstra, J. J. Bax, J. H. C. Reiber, and A. J. Scholte, "Automatic detection and quantification of the Agatston coronary artery calcium score on contrast computed tomography angiography," Int. J. Cardiovasc. Imaging **31**, 151–161 (2014).

[13]Y. Xie, M. D. Cham, C. Henschke, D. Yankelevitz, and A. P. Reeves, "Automated coronary artery calcification detection on low-dose chest CT images," Proc. SPIE **9035**, 90350F (2014).

[14]D. Eilot and R. Goldenberg, "Fully automatic model-based calcium segmentation and scoring in coronary CT angiography," Int. J. Comput. Assisted Radiol. Surg. **9**, 595–608 (2014).

[15]R. A. P. Takx, P. A. De Jong, T. Leiner, M. Oudkerk, H. J. De Koning, C. P. Mol, M. A. Viergever, and I. Išgum, "Automated coronary artery calcification scoring in non-gated chest CT: Agreement and reliability," PLoS One **9**, e91239 (2014).

[16]R. Shahzad, T. van Walsum, M. Schaap, A. Rossi, S. Klein, A. C. Weustink, P. J. de Feyter, L. J. van Vliet, and W. J. Niessen, "Vessel specific coronary artery calcium scoring: An automatic system," Acad. Radiol. **20**, 1–9 (2013).

[17]J. Wu, G. Ferns, J. Giles, and E. Lewis, "A fully automated multi-modal computer aided diagnosis approach to coronary calcium scoring of MSCT images," Proc. SPIE **8315**, 83152I (2012).

[18]B. A. Arnold, P. Xiang, M. J. Budoff, and S. S. Mao, "Very small calcifications are detected and scored in the coronary arteries from small voxel MDCT images using a new automated/calibrated scoring method with statistical and patient specific plaque definitions," Int. J. Cardiovasc. Imaging **28**, 1193–1204 (2012).

[19]S. Mittal, Y. Zheng, B. Georgescu, F. Vega-Higuera, S. K. Zhou, P. Meer, and D. Comaniciu, "Fast automatic detection of calcified coronary lesions in 3D cardiac CT images," in *Machine Learning in Medical Imaging*, Lecture Notes in Computer Science, edited by F. Wang, P. Yan, K. Suzuki, and D. Shen (Springer, Berlin, Heidelberg, 2010), pp. 1–9.

[20]U. Kurkure, D. R. Chittajallu, G. Brunner, Y. H. Le, and I. A. Kakadiaris, "A supervised classification-based method for coronary calcium detection in non-contrast CT," Int. J. Cardiovasc. Imaging **26**, 817–828 (2010).

[21]G. Brunner, D. R. Chittajallu, U. Kurkure, and I. A. Kakadiaris, "Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data," Int. J. Cardiovasc. Imaging **26**, 829–838 (2010).

[22]S. C. Saur, H. Alkadhi, L. Desbiolles, G. Szekely, and P. C. Cattin, "Automatic detection of calcified coronary plaques in computed tomography data sets," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*, Lecture Notes in Computer Science, edited by D. Metaxas, L. Axel, G. Fichtinger, and G. Szekely (Springer, Berlin, Heidelberg, 2008), pp. 170–177.

[23]I. Isgum, A. Rutten, M. Prokop, and B. van Ginneken, "Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease," Med. Phys. **34**, 1450–1461 (2007).

[24]S. Wesarg, M. F. Khan, and E. A. Firle, "Localizing calcifications in cardiac CT data sets using a new vessel segmentation approach," J. Digital Imaging **19**, 249–257 (2006).

[25]D. Dey, V. Y. Cheng, P. J. Slomka, R. Nakazato, A. Ramesh, S. Gurudevan, G. Germano, and D. S. Berman, "Automated 3-dimensional quantification of noncalcified and calcified coronary plaque from coronary CT angiography," J. Cardiovasc. Comput. Tomogr. **3**, 372–382 (2009).

[26]I. Isgum, M. Prokop, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Automatic coronary calcium scoring in low-dose chest computed tomography," IEEE Trans. Med. Imaging **31**, 2322–2334 (2012).

[27]http://orcascore.isi.uu.nl/.

[28]S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," IEEE Trans. Med. Imaging **29**, 196–205 (2010).

[29]A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention* (Springer, Berlin, Heidelberg, 1998), pp. 130–137.

[30]G. Yang, P. Kitslaar, M. Frenay, A. Broersen, M. J. Boogers, J. J. Bax, J. H. C. Reiber, and J. Dijkstra, "Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography," Int. J. Cardiovasc. Imaging **28**, 921–933 (2012).

[31]Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," IEEE Trans. Med. Imaging **27**, 1668–1681 (2008).

[32]Y. Zheng, H. Tek, and G. Funka-Lea, "Robust and accurate coronary artery centerline extraction in CTA by combining model-driven and data-driven approaches," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, Lecture Notes in Computer Science, edited by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Springer, Berlin, Heidelberg, 2013), pp. 74–81.

[33]S. L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," Data Min. Knowl. Discovery **1**, 317–328 (1997).

[34]A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, "Quantification of coronary artery calcium using ultrafast computed tomography," J. Am. Coll. Cardiol. **15**, 827–832 (1990).

[35]I. Zeb, N. Abbas, K. Nasir, and M. J. Budoff, "Coronary computed tomography as a cost-effective test strategy for coronary artery disease assessment: A systematic review," Atherosclerosis **234**, 426–435 (2014).

[36]G. Gitsioudis, W. Hosch, J. Iwan, A. Voss, E. Atsiatorme, N. P. Hofmann, S. J. Buss, S. Siebert, H.-U. Kauczor, and E. Giannitsis, "When do we really need coronary calcium scoring prior to contrast-enhanced coronary computed tomography angiography? Analysis by age, gender and coronary risk factors," PLoS One **9**, e92396 (2014).

[37]R. Tota-Maharaj, M. H. Al-Mallah, K. Nasir, W. T. Qureshi, R. Blumenthal, and M. J. Blaha, "Improving the relationship between coronary artery calcium score and coronary plaque burden: Addition of regional measures of coronary artery calcium distribution," Atherosclerosis **238**, 126–131 (2014).

[38]E. R. Brown, R. A. Kronmal, D. A. Bluemke, A. D. Guerci, J. J. Carr, J. Goldin, and R. Detrano, "Coronary calcium coverage score: Determination, correlates, and predictive accuracy in the multi-ethnic study of atherosclerosis," Radiology **247**, 669–675 (2008).

[39]M. G. Silverman, J. R. Harkness, R. Blankstein, M. J. Budoff, A. S. Agatston, J. J. Carr, J. A. Lima, R. S. Blumenthal, K. Nasir, and M. J. Blaha, "Baseline subclinical atherosclerosis burden and distribution are associated with frequency and mode of future coronary revascularization: Multi-ethnic study of atherosclerosis," JACC: Cardiovasc. Imaging **7**, 476–486 (2014).

[40]M. Schaap *et al.*, "Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms," Med. Image Anal. **13**, 701–714 (2009).

[41]C. H. McCollough, S. Ulzheimer, S. S. Halliburton, K. Shanneik, R. D. White, and W. A. Kalender, "Coronary artery calcium: A multi-institutional, multimanufacturer international standard for quantification at cardiac CT," Radiology **243**, 527–538 (2007).

[42]M. J. Willemink, R. Vliegenthart, R. A. P. Takx, T. Leiner, R. P. J. Budde, R. L. A. W. Bleys, M. Das, J. E. Wildberger, M. Prokop, and N. Buls, "Coronary artery calcification scoring with state-of-the-art CT scanners from different vendors has substantial effect on risk classification," Radiology **273**, 695–702 (2014).

[43]P. H. Joshi, M. J. Blaha, R. S. Blumenthal, R. Blankstein, and M. P. H. Khurram Nasir MD, "What is the role of calcium scoring in the age of coronary computed tomographic angiography?," J. Nucl. Cardiol. **19**, 1226–1235 (2012).