Impact of automatically detected motion artifacts on coronary calcium scoring in chest computed tomography

Jurica Šprem
Bob D. de Vos
Nikolas Lessmann
Pim A. de Jong
Max A. Viergever
Ivana Išgum

Impact of automatically detected motion artifacts on coronary calcium scoring in chest computed tomography

Jurica Šprem,^{a,*} Bob D. de Vos,^a Nikolas Lessmann,^a Pim A. de Jong,^b Max A. Viergever,^c and Ivana Išgum^a

^aUniversity Medical Center Utrecht, Image Sciences Institute, Utrecht, The Netherlands

^bUtrecht University and University Medical Center Utrecht, Department of Radiology, Utrecht, The Netherlands

^cUtrecht University and University Medical Center Utrecht, Image Sciences Institute, Utrecht, The Netherlands

Abstract. The amount of coronary artery calcification (CAC) quantified in computed tomography (CT) scans enables prediction of cardiovascular disease (CVD) risk. However, interscan variability of CAC quantification is high, especially in scans made without ECG synchronization. We propose a method for automatic detection of CACs that are severely affected by cardiac motion. Subsequently, we evaluate the impact of such CACs on CAC quantification and CVD risk determination. This study includes 1000 baseline and 585 one-year follow-up low-dose chest CTs from the National Lung Screening Trial. About 415 baseline scans are used to train and evaluate a convolutional neural network that identifies observer determined CACs affected by severe motion artifacts. Therefore, 585 paired scans acquired at baseline and follow-up were used to evaluate the impact of severe motion artifacts on CAC quantification and risk categorization. Based on the CAC amount, the scans were categorized into four standard CVD risk categories. The method identified CACs affected by severe motion artifacts with 85.2% accuracy. Moreover, reproducibility of CAC scores in scan pairs is higher in scans containing mostly CACs not affected by severe cardiac motion. Hence, the proposed method enables identification of scans affected by severe cardiac motion, where CAC quantification may not be reproducible. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.5.4.044007](https://doi.org/10.1117/1.JMI.5.4.044007)]

Keywords: calcium scoring; cardiac motion; chest computed tomography; convolutional neural network.

Paper 18131RR received Jun. 14, 2018; accepted for publication Nov. 16, 2018; published online Dec. 11, 2018.

1 Introduction

The amount of coronary artery calcification (CAC) is an independent and strong predictor of cardiovascular diseases (CVDs), such as myocardial infarction or sudden cardiac death. CAC is typically quantified in ECG-triggered cardiac computed tomography (CT) scans acquired with high in-plane image resolution. Nevertheless, recent research shows that CAC burden can also be determined in any CT scan that visualizes the heart.^{1–4}

With the implementation of lung cancer screening in the USA⁵ and possibly also in Europe,^{6–8} the potential of CAC quantification, i.e. CAC scoring, in chest CT and thereby identification of subjects at risk of CVD in the screening population, has become a topic of intensive research.^{2,3,9} The US Preventive Services Task Force recommends annual screening for lung cancer with low-dose CT in adults aged 55 to 80 years, who have a smoking history of at least 30-pack years and currently smoke or have quit smoking within the past 15 years.¹⁰ This may result in acquisition of several millions of low-dose CT scans annually. Furthermore, a joint guideline from the Society of Cardiovascular Computed Tomography and the Society of Thoracic Radiology endorses reporting the CAC burden on all noncontrast CT examinations.¹¹ Routine CAC reporting could provide the opportunity to follow CAC progression over time.

In order to follow a subject's CAC score over time, a reproducible scoring method is required. However, several

studies have shown that CAC scoring has limited reproducibility even when performed in cardiac CT scans that are synchronized with cardiac motion.^{12–15} Detrano et al.¹⁵ have shown that 28% of cardiac scans obtained from the Multiethnic Study of Atherosclerosis using multiple detector computed tomography (MDCT) had excessive motion artifacts. Elss et al.¹⁶ investigated identifying motion artifacts on artificially generated motion in cardiac CT scans. In chest CT scans, reproducibility is additionally compromised, by the lack of ECG synchronization leading to cardiac motion artifacts, by low image resolution leading to partial volume effects, and possibly by suboptimal image reconstruction kernels resulting in high levels of image noise.^{17–20}

We propose a method for automatic identification of CACs that are severely affected by cardiac motion. Subsequently, we evaluate the impact of such CACs on their quantification and CVD risk determination. This study includes low-dose chest CTs from the National Lung Screening Trial (NLST) acquired at baseline and 1-year follow-up.²¹ Baseline scans are used to train and evaluate a convolutional neural network (CNN) that identifies CACs affected by severe motion artifacts. Thereafter, an additional set of paired scans acquired at baseline and follow-up were used to evaluate the impact of motion artifacts affecting CACs on CAC scoring expressed with Agatston score²² and CVD risk categorization. This paper extends our preliminary work described in Šprem et al.²³ that outlines the identification of CACs severely affected by cardiac motion.

*Address all correspondence to Jurica Šprem, E-mail: j.sprem@umcutrecht.nl

2 Data

Data used in this study included 1585 scans of 1000 subjects participating in the NLST. Participants enrolled in the NLST were between 55 and 74 years of age and were either current or former smokers.²¹ Each of the 33 screening centers received institutional review board approval before the onset of recruitment and all participants signed a written consent regarding the use of data.²⁴

Scans were obtained using different multidetector CT scanners from four major vendors (General Electric, Philips, Siemens, and Toshiba). Scans were made using low irradiation dose, in a single maximal breath hold, without the use of intravenous contrast medium, and without specified use of ECG synchronization.²⁴ The tube voltage was set to 120 or 140 kVp, depending on the weight of the participant, and the tube current ranged from 40 to 80 mAs. In-plane resolution was 512×512 voxels with sizes ranging from 0.47 to 0.89 mm with 2- to 3.2-mm slice thickness and 1.25- to 2.5-mm slice spacing. Prior to analysis, the scans were reconstructed to 3-mm slice thickness with 1.5-mm slice spacing to perform CAC quantification with recommended settings.¹⁹ In all scans, automatic segmentation of CAC lesions was performed using an algorithm developed by Lessmann et al.²⁵ and previously validated on NLST scans.

The scans were divided into two subsets. The first subset contained baseline scans of 415 subjects. 61% of scans were acquired with GE, 32% with Siemens, 3.5% from Philips, and 3.5% with Toshiba scanner. In these scans, each 2-D CAC was labeled as severely affected by cardiac motion (motion affected CAC) or not (motion-free CAC) by visual evaluation. We defined CAC to be affected by severe motion artifacts if it had crescent, star-like or smeared line shapes (Fig. 1). These scans were used for method development and evaluation. The second subset contained 585 scan pairs obtained (49.5% with

GE, 34% with Siemens, 8.5% from Philips, and 8% from Toshiba scanner) at baseline and 1-year follow-up. From this subset, 41 scan pairs were excluded because of zero CAC score in at least one of the scans as motion detection requires detected calcium. In addition, 28 scan pairs were excluded since at least one of the scans contained extreme levels of image noise preventing CAC scoring. The remaining 516 scan pairs (1032 CT scans) were used to investigate the impact of motion artifacts on CAC score and risk categorization in each scan. The scans were categorized in four risk categories based on the CAC score (I: < 10 , II: 10 to 100, III: 100 to 400, IV: > 400).

3 Methods

3.1 Classification of 2-D CAC Lesions According to Cardiac Motion

To discriminate CACs severely affected by cardiac motion from those that are not, a CNN is used. CNNs have proven to be extremely powerful in image classification without the need for handcrafting image features. To identify CACs affected by cardiac motion artifacts, we investigate two CNN architectures, CNN-1 and CNN-2 with a similar number of weights. Both networks take 2-D patches from axial plane centered on a CAC lesion as input. CNN-1 has a narrow design having the same number of filters in all convolutional layers, whereas CNN-2 has a trapezoidal design inspired by previously described VGGnet architecture.²⁶ The input patch containing CAC is processed by three convolutional layers using 3×3 kernels, each followed by subsampling layers of 2×2 maxpooling. CNN-1 has 32 filters in each convolutional layer, whereas CNN-2 has 16, 32, and 64 filters, consecutively (Table 1). The last maxpooling layer links to two consecutive fully connected

Fig. 1 (a) Examples of the same CACs found in CT slices from baseline and 1-year follow-up scans. Blue squares depict CACs not severely affected by cardiac motion. Red squares depict CACs severely affected by cardiac motion. (b) Examples of one 3-D CAC in the right coronary artery visible in multiple consecutive axial (2-D) slices with their corresponding labels.

Table 1 Characteristics of the two CNN architectures.

CNN-1	CNN-2
32 - conv3 × 3	16 - conv3 × 3
max-pool2 × 2	max-pool2 × 2
32 - conv3 × 3	32 - conv3 × 3
max-pool2 × 2	max-pool2 × 2
32 - conv3 × 3	64 - conv3 × 3
max-pool2 × 2	max-pool2 × 2
16 - Fully connected	
4 - Fully connected	
2 - Softmax	

layers of 16 and 4 nodes, followed by a softmax layer with two units providing the final classification label; see Fig. 2 for a schematic overview.

We evaluated different input patch sizes: 48×48 , 64×64 , and 92×92 pixels. The smallest patch size (48×48) was selected as most of the 2-D CAC lesions are fully contained in it. Given the fact that the coronaries mostly run on the heart surface, the largest patch size (92×92) was selected to cover a large part of the heart while limiting the analysis of extracardiac regions (e.g., the lungs). Using the developed CNN, each 2-D CAC lesion is classified as either motion affected or motion-free lesion (see Sec. 2).

3.2 Impact of Motion Artifacts on CAC Score and CVD Risk Categorization

The CVD risk of a subject is determined by the total CAC score in a scan. A scan may contain a single large CAC lesion being severely affected by motion leading to a low CAC scoring reproducibility. However, a scan may have also several smaller CAC lesions affected by motion that do not appreciably affect scoring reproducibility. To identify scans affected by cardiac motion that result in low CAC scoring reproducibility, we investigated two criteria: the percentage of motion affected 2-D CAC lesions in a scan and the percentage of motion affected CAC volumes in the scan. Based on the percentage of motion affected CACs

(volumes), a scan is classified as motion affected or motion-free scan. Various thresholds on the level of motion affected CACs were investigated for both criteria. Additionally, we investigated how the number of motion-free scans that are considered to be suitable for CAC scoring changes depending on different threshold. Hence, a motion-free scan can still contain CACs affected by motion artifacts but below the defined threshold.

To investigate reproducibility of CAC scores, we compared the CAC scores and CVD risk categories in the baseline and follow-up scans of the same subject. For this purpose, we defined two groups of scans. The motion-free group contains pairs of scans that are both classified as motion-free scan. The motion group contains pairs of scans, where at least one of the scans is classified as motion affected scan. To evaluate the impact of cardiac motion affecting CACs on the interscan agreement of CAC scores, the intraclass correlation coefficient (ICC) for the scan pairs was computed. This was done for motion and motion-free groups separately. Also, the percentage of scans suitable for CAC scoring (motion-free scans) was determined.

To assess the impact of motion artifacts on CVD risk categorization, each scan was assigned to one of the four common risk categories based on the Agatston CAC score: <10 , 10 to 100, 100 to 400, >400 . Note that pairs of scans where at least one of the scans had zero CAC score were excluded from the analysis. Thereafter, a nonparametric Wilcoxon signed rank test was used to determine whether there are significant differences in risk category assignment between the scan pairs. The confusion matrix showing risk categorization between scan pairs for the motion-free group and the motion group was computed.

4 Experiments and Results

4.1 Classification of CAC Lesions According to Cardiac Motion

The CNNs were trained and validated on 352 randomly selected images containing at least one CAC resulting in 14,504 extracted 2-D patches containing CAC. Randomly selected 80% of images were used for training and 20% for validation (5215 motion affected and 9289 motion-free). The final performance was evaluated using a separate test set containing 63 images with 1974 extracted patches (591 motion affected and 1383 motion-free). The CNNs were trained with a mini-batch size of 128 samples to minimize the loss function, optimizing weights using the Adam optimizer²⁷ with a learning rate of

Fig. 2 Illustration of the CNN-1 architecture with input patches. The network contains three convolutional layers with $32 \times 3 \times 3$ kernels, each followed by a max-pooling 2×2 subsampling layer. The last max-pooling layer links to two consecutive fully connected layers with 16 and 4 nodes, followed by a softmax layer with two outputs.

Table 2 Classification performance of the CNN-1 and CNN-2 presented with accuracy and AUC based on different input patch sizes.

Patch size (pixel)	CNN-1		CNN-2	
	Accuracy	AUC	Accuracy	AUC
48 × 48	85.2%	0.90	79.7%	0.86
64 × 64	81.2%	0.87	78.9%	0.85
92 × 92	80.9%	0.87	77.0%	0.82

0.0001. A random dropout²⁸ of 50% was applied between fully connected layers to prevent overfitting.

We compared classification performance of the CNNs using several patch sizes. As shown in Table 2, CNN-1 had consistently higher performance than CNN-2. Both CNN-1 and CNN-2 achieved highest performance with patches of size 48 × 48 pixels. Hence, this size was used in further analysis. CNN-1 achieved an accuracy of 85.2% (76.2% sensitivity, 88.8% specificity) and area under the receiver operating characteristic curve (AUC) of 0.90. The accuracy of CNN-1 on scans per scanning vendor was: GE 83.4%, Siemens 88.7%, Philips 87.4%, and Toshiba 90.6%. Examples of correctly classified patches without and with motion artifacts are shown in Fig. 3.

Visual inspection of misclassified CACs showed that most of the errors occurred because of mild motion artifacts [Figs. 4(a) and 4(c)], whereas the rest of the misclassified patches contained a large amount of noise and CACs affected by motion artifacts with CAC components having intensities below the 130 HU detection threshold [Figs. 4(b) and 4(d)].

4.2 Impact of Motion Artifacts on CAC Score and Risk Categorization

To investigate the impact of different amounts of motion artifacts on the CAC score, scans were separated into motion and motion-free groups based on the percentage of motion affected CACs or the percentage of motion affected CACs volume (Sec. 3.2). ICC was calculated between Agatston CAC scores for the pairs of baseline and follow-up scans (Table 3).

Depending on the threshold on the maximum percentage of the allowed motion CACs in a scan, the number of subjects and the interscan agreement of CAC scores in the groups vary. To determine the cause of the risk category change, we visually inspected the 147 subjects that have had risk category change between the baseline and follow-up scans. Among them, 96 subjects had an increase and 51 subjects had a decrease in the CVD risk category. The analysis shows that CAC progression is often not related by image acquisition or reconstruction parameters, whereas CAC regression can often be attributed to cardiac motion or other factors causing image quality degradation.

Fig. 3 Examples of correctly classified patches containing CACs using CNN-1. Top: CACs not affected by motion artifacts; bottom: CACs affected by severe motion artifacts.

Fig. 4 Examples of incorrectly classified patches containing CACs: (a) CAC incorrectly classified as affected by motion artifact; (b) noisy patch containing CAC labeled as not affected but classified as affected by motion artifact; (c) CAC incorrectly classified as not affected by motion artifact; and (d) noisy patch misclassified as not affected by motion artifact.

Table 3 Quantitative analysis of the influence of motion artifacts on the reproducibility of determined CAC scores between the pairs of baseline and 1-year follow-up scans. The results are ordered based on the maximum percentage of motion artifacts within each of the paired scans. Two different criteria were investigated: the percentage of 2-D CACs and the volume percentage of CACs affected by severe motion artifacts within a scan.

Based on the threshold for the percentage of motion affected 2-D CACs in a scan						
Percentage of motion artifacts in a scan	Motion-free group			Motion group		
	Percentage of subjects	ICC	CI _{95%}	Percentage of subjects	ICC	CI _{95%}
—	100%	0.909	[0.900–0.924]	0%	—	—
<90%	98.4%	0.908	[0.889–0.924]	1.6%	0.719	[0.060–0.938]
<80%	96.1%	0.919	[0.902–0.933]	3.9%	0.757	[0.490–0.895]
<70%	90.9%	0.931	[0.917–0.943]	9.1%	0.804	[0.667–0.887]
<60%	81.8%	0.940	[0.926–0.951]	18.2%	0.804	[0.716–0.867]
<50%	72.1%	0.942	[0.928–0.954]	27.9%	0.861	[0.808–0.899]
<40%	59.7%	0.932	[0.913–0.946]	40.3%	0.891	[0.857–0.917]
<30%	51.4%	0.937	[0.918–0.951]	48.6%	0.892	[0.861–0.916]
<20%	39.9%	0.925	[0.902–0.943]	60.1%	0.900	[0.872–0.921]
<10%	27.7%	0.927	[0.899–0.947]	72.3%	0.901	[0.877–0.921]
Based on the threshold for the percentage of motion affected CACs volume in a scan						
Percentage of motion artifacts in a scan	Motion-free group			Motion group		
	Percentage of subjects	ICC	CI _{95%}	Percentage of subjects	ICC	CI _{95%}
—	100%	0.909	[0.900–0.924]	0%	—	—
<90%	92.9%	0.915	[0.897–0.930]	7.1%	0.862	[0.748–0.927]
<80%	82.8%	0.941	[0.927–0.952]	17.2%	0.793	[0.697–0.861]
<70%	72.6%	0.948	[0.934–0.959]	27.4%	0.829	[0.768–0.875]
<60%	63.1%	0.931	[0.912–0.945]	36.9%	0.887	[0.850–0.915]
<50%	53.2%	0.931	[0.912–0.946]	46.8%	0.892	[0.859–0.910]
<40%	45.6%	0.923	[0.902–0.940]	54.4%	0.897	[0.867–0.920]
<30%	38.1%	0.939	[0.920–0.954]	61.9%	0.896	[0.875–0.920]
<20%	31.3%	0.925	[0.898–0.945]	68.7%	0.901	[0.871–0.920]
<10%	25.4%	0.806	[0.736–0.859]	74.6%	0.903	[0.879–0.920]

As shown in Fig. 5, a mild threshold that allows many motion affected CACs in a scan, e.g., 60%–90%, defines many scans as motion-free. However, a large portion of subject within motion-free group changed their risk category due to CAC being affected by motion artifacts. By contrast, a strict threshold allowing only a few motion affected CACs in a scan, e.g., 10%–40%, defines a small number of scans as motion-free. Furthermore, subjects from motion group defined by strict threshold show substantial risk category change caused by

CACs being affected by severe motion artifacts. However, note that regardless of the threshold value, a better interscan agreement is noted in the motion-free than in the motion group for both criteria regarding presence of motion affected CACs per scan. This is not valid only for 10% threshold defined on the percentage of motion affected CACs volume. As motion artifacts are not the only factor negatively influencing interscan agreement, a strict threshold does not necessarily give a better interscan agreement in the motion-free group than using a mild

Fig. 5 Subjects with increased or decreased risk category between baseline and follow-up scan. Each subject is classified in motion or motion-free group, based on a threshold applied on the maximum percentage of 2-D CACs affected by motion artifacts. On y-axis is the number of subjects and on x-axis is the applied threshold defining the motion or motion-free group. The change in risk category is due to (1) motion artifacts, (2) other image degradation factors (image noise or larger in-plane voxel size leading to partial volume effect), or (3) no visible reason. (a) Risk category increase (96 subjects) and (b) risk category decrease (51 subjects).

threshold (Fig. 5). The used threshold may depend on a specific application.

We have further focused our analysis on a mild (30% of scan's CACs affected by motion) and a strict (70% of scan's CACs affected by motion) threshold. With these thresholds, the number of paired scans in both groups is comparable. After separating scan pairs into motion and motion-free groups based on the two thresholds, the changes in CAC risk

categorization were investigated (Table 4). Regardless of the used threshold for (volume) CAC percentage affected by motion artifacts, results in the motion-free group indicate possible progression with significant differences of risk category assignments between baseline and follow-up scans.

In Table 5, we further detail the changes in subjects' risk categorization between the baseline and follow-up scans for the motion-free group, the motion group, and a subgroup of

Table 4 Qualitative analysis of changes in risk categorization between subjects' baseline and follow-up scans in the motion affected and motion-free groups. Groups are separated by using a threshold either for the percentage or for the volume percentage of CACs affected by motion artifacts in a scan. Numbers of subjects with changed risk category between baseline and follow-up scans are listed as increased (higher risk) and decreased (lower risk), respectively, while the numbers of subjects with unchanged risk category are listed under tied. Significance of changes of category within each group (significant if $p < 0.05$) was determined using the Wilcoxon signed ranked nonparametric test.

Threshold: percentage of motion affected 2-D CACs in a scan						
Group	Threshold	No. of subjects	Increased	Decreased	Tied	<i>p</i> -value
Motion-free	70%	469 (90.9%)	87	47	335	0.001
Motion		47 (9.1%)	9	4	34	0.166
Motion-free	30%	265 (51.4%)	63	27	175	0.0002
Motion		251(49.6%)	33	24	194	0.233
Threshold: percentage of motion affected CACs volume in a scan						
Group	Threshold	No. of subjects	Increased	Decreased	Tied	<i>p</i> -value
Motion-free	70%	378 (73.3%)	76	35	267	0.0001
Motion		138 (66.7%)	20	16	102	0.505
Motion-free	30%	201 (38.9%)	54	21	126	0.0002
Motion		315 (61.1%)	42	30	243	0.157

the latter group (both scans classified as motion affected) defined using percentage of motion affected CACs volume. The motion affected subgroup indicates similar progression and regression of subjects' risk categories with strict threshold (30%), however, for the mild threshold (70%), that does not seem to be the case. The motion-free group indicates progression for both strict and mild threshold. With strict threshold, the CAC categories within the motion-free group show higher CAC progression, as images contain fewer CACs with severe motion artifacts, which introduce errors to CAC categorization. We investigated using Cohen's kappa value for a measurement of agreement between the two categorical variables weighting the extent of disagreement. Linearly weighted kappa for mild threshold ($<70\%$) is 0.73 and 0.71 in motion-free and motion group, respectively. For strict threshold ($<30\%$), linearly weighted kappa is 0.60 and 0.76 for motion-free and motion group, respectively. The lower kappa value for strict threshold and the results listed in Table 4 in the motion-free group indicates more subjects with CAC progression than the group with a mild threshold. Note that the numbers of subjects within the groups with strict and mild thresholds are different. Similar trends are observed for the same thresholds on the percentage of 2-D CACs affected by motion artifacts.

5 Discussion

Here, we have proposed and evaluated a CNN-based method for automatic detection of cardiac motion artifacts that severely affect CACs in chest CT scans. Depending on the extent of motion artifacts affecting CACs within a scan, each is classified as motion affected or motion-free scan. Paired baseline and 1-year follow-up CT scans are assigned to one of two groups: motion group if any of the paired scans is severely affected by cardiac motion or motion-free group otherwise. A high number

of CACs with severe motion artifacts in a pair negatively impacts the interscan agreement of the CAC score and, consequently, cardiovascular risk categorization. While CAC scores in scan pairs with a low number of CACs affected by motion artifacts may indicate possible progression of the CAC score in 1-year follow-up time, such a prediction cannot reliably be made from scan pairs that are heavily affected by motion artifacts.

To the best of our knowledge, the presented method is the first proposing automatic image analysis approach for the detection of motion artifacts affecting CACs in chest CT scans used to evaluate their impact on the reproducibility of CAC scoring. The presented CNN approach removes the need for manual intervention utilized by other studies^{15,29} or additional information acquired during the scan, such as sinogram data.³⁰ The CNN architecture and patch size were tuned to this specific application, but possibly other architectures would be suitable for the task as well.

The motion artifacts are mainly manifested in the axial plane owing to the nature of cardiac motion in combination with CT's helical scanning and highest resolution in axial plane. Furthermore, smaller lesions are visible in a single slice only, but large lesions may be present in multiple slices. In the latter case, the lesions may not be equally affected by cardiac motion in each slice. Therefore, to evaluate the extent of motion artifacts affecting CACs within a scan, 2-D patches centered around CAC lesions in axial slice were used for classification of motion artifacts. Subjects with high CAC score (high CVD risk category) often have more calcifications and they are more often large. Hence, the chance that such subjects have calcification affected by severe motion is larger than in subjects with few small calcifications.

In this study, the extent of motion artifacts affecting CACs within a scan is expressed by the percentage or by the volume

Table 5 Confusion matrices for the motion-free group, the motion group, and a subgroup of the motion group (both scans classified as motion affected), separated based on thresholds of 30% (with low number of motion artifacts per scan) and 70% (with high number of motion artifacts per scan) of CAC volumes affected by motion artifacts. The confusion matrices present percentages of subjects within four risk categories based on the CAC score (I: <10, II: 10 to 100, III: 100 to 400, IV: <400).

< 30% of CAC volumes affected by motion artifacts											
Motion-free—201 subjects						Motion—315 subjects					
		Follow-up						Follow-up			
		I	II	III	IV			I	II	III	IV
Baseline	I	14.9	14.4	0.5	0.0	Baseline	I	2.2	2.5	0.0	0.0
	II	5.5	26.9	10.4	0.5		II	1.6	15.2	4.44	0.0
	III	0.5	3.5	12.9	1.0		III	0.0	4.1	26.9	6.3
	IV	0.0	0.0	1.0	8.0		IV	0.0	0.0	3.8	32.7
						Motion (subgroup)—180 subjects					
		Follow-up						Follow-up			
		I	II	III	IV			I	II	III	IV
Baseline	I	0.50	2.2	0.0	0.0	Baseline	I	0.50	2.2	0.0	0.0
	II	0.0	11.1	3.3	0.0		II	0.0	11.1	3.3	0.0
	III	0.0	3.3	27.8	6.1		III	0.0	3.3	27.8	6.1
	IV	0.0	0.0	4.4	40.6		IV	0.0	0.0	4.4	40.6
<70% of CAC volumes affected by motion artifacts											
Motion-free—378 subjects						Motion—138 subjects					
		Follow-up						Follow-up			
		I	II	III	IV			I	II	III	IV
Baseline	I	9.3	7.9	0.5	0.0	fBaseline	I	1.5	4.4	0.0	0.0
	II	3.7	22.8	7.9	0.3		II	1.5	11.6	3.6	0.0
	III	0.3	3.4	18.8	3.4		III	0.0	5.1	29.0	6.5
	IV	0.0	0.0	1.9	19.8		IV	0.0	0.0	5.1	31.9
						Motion (subgroup)—50 subjects					
		Follow-up						Follow-up			
		I	II	III	IV			I	II	III	IV
Baseline	I	0.0	6.0	0.0	0.0	Baseline	I	0.0	6.0	0.0	0.0
	II	0.0	6.0	0.0	0.0		II	0.0	6.0	0.0	0.0
	III	0.0	0.0	22.0	10.0		III	0.0	0.0	22.0	10.0
	IV	0.0	0.0	4.0	52.0		IV	0.0	0.0	4.0	52.0

percentage of CAC lesions affected by motion artifacts. Following other studies,^{15,29} binary selection of scans (according to no motion to the presence of severe motion artifacts) was used to identify subjects having scans with CAC scores highly compromised by motion artifacts. Regardless of the used criterion, a substantial impact of motion artifacts is observed. We also investigated several thresholds to classify scans as being severely affected by motion artifacts. Depending on the strictness of the threshold, the interscan agreement (ICC) of CAC scores and the number of scans in the group defined as suitable for CAC scoring (motion-free) vary. The motion-free group defined by a strict threshold shows a high ratio of subjects having CAC regression caused by having CACs affected by motion artifacts. However, for both mild and strict thresholds using both criteria, the interscan agreement of CAC scoring is higher in the motion-free group than in the motion group. Given that the number of subjects defined suitable for CAC scoring (motion-free) as well as the corresponding CAC score reproducibility varies, the acceptable amount of severe motion artifacts within a scan should be tuned to the objective of the study at hand, as motion artifacts are not the only factor influencing scoring reproducibility.

In this paper, we aimed to investigate the impact of severe motion artifacts on CAC scoring and CVD risk categorization. In future work, determining a threshold value for tolerated level of acceptable motion artifacts needs to be investigated. This value may depend on the specific application. A study focusing on CAC progression, a high CAC score reliability, is favored over a high number of included scans. Possibly, an epidemiology study, where only CVD risk categorization is important, may favor including a large number of scans over very strict threshold on the motion artifact. As shown in this study, removing the scans most severely affected by severe motion artifacts already gives clearer insight into the impact motion has in a study by negatively influencing CAC scoring. Inasmuch as the CAC score is an important predictor of CVD events, the impact of severe motion artifacts on cardiovascular risk categorization was investigated. Scans within the motion-free group demonstrate higher CAC score reproducibility and significant difference between assigned risk categories based on baseline and 1-year follow-up scans indicating CAC progression. An observed higher risk category in the follow-up scan than in the baseline scan could imply that CAC progression can be established over a short period of time. In the motion group, which is severely affected by motion artifacts, there is no significant difference of risk categories between baseline and follow-up scans while the number of subjects showing both progression and regression of risk categories is similar. As regression is not expected and is considered to be caused by errors, this indicates the high impact of severe motion artifacts on the interscan variability of CAC scoring.

In this study, we have demonstrated the high impact of severe motion artifacts on CAC score and CVD risk categorization in chest CT scans. Chest CT scans are primarily made to image lungs, which implies they are suboptimal for CAC scoring. Cardiac motion can decrease CAC intensities below the level of detection, which may imply that only few, or even no, voxels of a CAC lesion are observed, which compromises the detection of motion artifacts in these cases. Besides motion artifacts, there are other factors influencing CAC score reproducibility, such as partial volume effect or high image noise caused by reconstruction parameters interfering with CAC scoring. This is the reason why with increasing the threshold for the percentage

of motion artifacts affecting CACs per scan is not necessarily resulting in increasing the ICC score. It might hence be beneficial to research lower CAC intensity detection levels and partial volume correction methods. The number of scans used for the CNN training was not balanced with respect to the number of scans acquired with different scanner vendors. Therefore, to make a solid conclusion regarding possible difference between the scanning vendors, a dedicated vendor specific study with a comparable number of scans from all vendors is required. In future work, it might be interesting to grade motion artifacts affecting calcium scoring on multiple levels (e.g., no motion, moderate motion, severe motion), but visually defining the reference standard for this task in a reproducible way is highly challenging. Furthermore, as in this study, only a single rater for severe motion artifacts was used, it would be constructive to extend the present study with interobserver agreement between multiple raters.

Even though chest CT scans differ substantially from cardiac CT scans that focus on imaging the heart, other CT scans that show the heart can be used for CAC scoring.^{3,18,20} Such studies may contain large numbers of scans and multiple follow-up scans that are acquired without the use of beta blockers or ECG-triggering. The use of an automatic tool for motion detection in such studies selecting the scans suitable for CAC scoring would allow investigating quantification, progression and changes of CACs more accurately. Furthermore, to the best of our knowledge, a study investigating CAC progression using chest CT scans is not yet performed. A dedicated study for CAC score progression in chest CT scans should be conducted taking in consideration, the presented impact motion artifacts have on CAC quantification.

With CAC score being a strong predictor of cardiovascular events, a CVD event prediction model could potentially benefit from additional information regarding the presence of severe motion artifacts in a scan and CAC scores corrected for partial volume. In our future work, we will investigate the predictive power of a such model.

6 Conclusion

The impact of motion artifacts on coronary artery calcium scoring and cardiovascular risk categorization in chest CTs was investigated. The proposed method is able to distinguish scans severely affected by cardiac motion from scans suitable for CAC scoring. Scans containing CACs severely affected by motion artifacts demonstrate lower CAC score reproducibility and cardiovascular risk categorization. Scans without an appreciable amount of motion artifacts allow better tracking of CAC score progression over time. Accordingly, the proposed method could potentially improve the accuracy and reliability of coronary artery calcium scoring and risk prediction from chest CT screening scans.

Disclosures

I. Isgum and M. A. Viergever have received an institutional research grant from PIE Medical Imaging B.V.; a research grant by the Netherlands Organization for Health Research and Development (ZonMw) with participation of PIE Medical Imaging B.V.; a research grant by Dutch Technology Foundation (TTW Perspectief) in the framework of the deep learning in Medical Image Analysis program (DLMedIA), with participation of Philips Healthcare and Pie Medical Imaging B.V.; and a research grant by Dutch Technology

Foundation (TTW Perspectief) in the framework of the Population Imaging Genetics (ImaGene) program within Project 12726 awarded with participation of Pie Medical Imaging B.V. and 3mensio Medical Imaging, which funded this research. The Department of Radiology of the University Medical Center Utrecht receives research support from Philips Healthcare.

Acknowledgments

This study was financially supported by the Netherlands Organization for Scientific Research (NWO)/Foundation for Applied and Engineering Sciences (STW); Project 12726. The authors thank the National Cancer Institute for access to NCI's data collected by the NLST. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

References

1. P. C. Jacobs et al., "Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer," *Am. J. Roentgenol.* **198**(3), 505–511 (2012).
2. O. M. Mets et al., "Lung cancer screening CT-based prediction of cardiovascular events," *JACC Cardiovasc. Imaging* **6**, 899–907 (2013).
3. C. Chiles et al., "Association of coronary artery calcification and mortality in the national lung screening trial: a comparison of three scoring methods," *Radiology* **276**(1), 82–90 (2015).
4. I. Isgum et al., "Automatic determination of cardiovascular risk by CT attenuation correction maps in Rb-82 PET/CT," *J. Nucl. Cardiol.* 1–10 (2017).
5. V. A. Moyer and United States Preventive Services Task Force, "Screening for lung cancer: US preventive services task force recommendation statement," *Ann. Intern. Med.* **160**(5), 330–338 (2014).
6. J. K. Field et al., "European randomized lung cancer screening trials: post NLST," *J. Surg. Oncol.* **108**(5), 280–286 (2013).
7. D. Shlomi et al., "Screening for lung cancer: time for large-scale screening by chest computed tomography," *Eur. Respir. J.* **44**(1), 217–238 (2014).
8. H.-U. Kauczor et al., "ESR/ERS white paper on lung cancer screening," *Eur. Radiol.* **25**, 2519–2531 (2015).
9. G. Chassagnon and M.-P. Revel, "Lung cancer screening: current status and perspective," *Diagn. Interventional Imaging* **97**(10), 949–953 (2016).
10. United States Preventive Services Task Force, "Final recommendation statement: lung cancer: screening," US Preventive Services Task Force, Rockville, Maryland, 2014, www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/lung-cancer-screening (30 December 2016).
11. H. S. Hecht 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.* **11**(1), 74–84 (2017).
12. S. Achenbach et al., "In-plane coronary arterial motion velocity: measurement with electron-beam CT," *Radiology* **216**(2), 457–463 (2000).
13. S. Mao et al., "Effect of electrocardiogram triggering on reproducibility of coronary artery calcium scoring," *Radiology* **220**(3), 707–711 (2001).
14. J. Horiguchi et al., "The impact of motion artifacts on the reproducibility of repeated coronary artery calcium measurements," *Eur. Radiol.* **17**(1), 81–86 (2007).
15. R. C. Detrano et al., "Coronary calcium measurements: effect of CT scanner type and calcium measure on rescans reproducibility—MESA study," *Radiology* **236**(2), 477–484 (2005).
16. T. Elss et al., "Deep-learning-based CT motion artifact recognition in coronary arteries," *Proc. SPIE* **10574**, 1057416 (2018).
17. A. Rutten, I. Isgum, and M. Prokop, "Coronary calcification: effect of small variation of scan starting position on Agatston, volume, and mass scores," *Radiology* **246**(1), 90–98 (2008).
18. P. C. A. Jacobs et al., "Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement," *Am. J. Roentgenol.* **194**(5), 1244–1249 (2010).
19. A. Rutten, I. Isgum, and M. Prokop, "Calcium scoring with prospectively ECG-triggered CT: using overlapping datasets generated with MPR decreases inter-scan variability," *Eur. J. Radiol.* **80**(1), 83–88 (2011).
20. S. A. M. Gernaat et al., "Automatic coronary artery calcium scoring on radiotherapy planning CT scans of breast cancer patients: reproducibility and association with traditional cardiovascular risk factors," *PLoS One* **11**, e0167925 (2016).
21. The National Lung Screening Trial Research Team, "The National Lung Screening Trial: overview and study design," *Radiology* **258**(1), 243–253 (2011).
22. A. S. Agatston et al., "Quantification of coronary artery calcium using ultrafast computed tomography," *J. Am. Coll. Cardiol.* **15**, 827–832 (1990).
23. J. Šprem et al., "Classification of coronary artery calcifications according to motion artifacts in chest CT using a convolutional neural network," *Proc. SPIE* **10133**, 101330R (2017).
24. D. R. Aberle et al., "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N. Engl. J. Med.* **365**(5), 395–409 (2011).
25. N. Lessmann et al., "Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions," *IEEE Trans. Med. Imaging* **37**(2), 615–625 (2018).
26. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
27. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
28. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
29. A. R. Paixao et al., "Defining coronary artery calcium concordance and repeatability—implications for development and change: the Dallas heart study," *J. Cardiovasc. Comput. Tomogr.* **11**(5), 347–353 (2017).
30. A. Schindler et al., "Iterative image reconstruction techniques for CT coronary artery calcium quantification: comparison with traditional filtered back projection in vitro and in vivo," *Radiology* **270**(2), 387–393 (2014).

Jurica Šprem obtained his MSc degree in information and communication technology from the University of Zagreb (2014). In 2014, he started as a PhD candidate at the Image Sciences Institute in UMC Utrecht, where his main area of research is machine learning in medical imaging with a focus in decreasing interscan variabilities, using of calcium scoring of plaque in coronary arteries and different machine learning approaches.

Bob D. de Vos finished his master's degree in biomedical engineering from the University of Groningen (RUG) (2012). During his study, he focused on clinical physics and became interested in medical image processing. He is currently working as a PhD-candidate focusing on machine learning with an emphasis on finding biomarkers in CT lung cancer screening scans to predict cardiovascular risk.

Nikolas Lessmann received his MSc degree in biomedical engineering from the University of Lübeck, Germany, in 2014. He is currently a PhD candidate at the Image Sciences Institute, University Medical Center Utrecht, the Netherlands, working on automatic calcium scoring and early detection of osteoporosis in lung cancer screening with low-dose chest CT.

Pim A. de Jong has been working as a radiologist at University Medical Center (UMC) Utrecht since completing the radiologist training at Meander Medical Center and at UMC Utrecht. He studied medicine at Erasmus University Rotterdam and also carried out PhD research on CT of the lungs of cystic fibrosis patients. At the University of Utrecht, he completed a master's degree in clinical epidemiology.

Max A. Viergever received his MSc degree in applied mathematics (1972) and his DSc degree in technical sciences (1980) from Delft University of Technology. He was appointed professor and head of the Department of Medical Imaging at Utrecht University in 1988; he retired in 2018. He is an elected fellow of IAPR, IEEE, IOP,

and MICCAI. He has supervised >150 PhD theses. From 2001 to 2008, he was the editor-in-chief of *IEEE Transactions on Medical Imaging*.

Ivana Išgum graduated in mathematics at the University of Zagreb (1999) and obtained a PhD degree at University Medical Center

(UMC) Utrecht (2007). She then worked for a year as a postdoc at the Laboratory for Clinical and Experimental Image Processing in Leiden University Medical Center. She is currently an associate professor at UMC Utrecht, where she is leading the Quantitative Medical Image Analysis (QIA) Group at the Image Sciences Institute.