Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT

van Assen, Marly; Martin, Simon S.; Varga-Szemes, Akos; Rapaka, Saikiran; Cimen, Serkan; Sharma, Puneet; Sahbaee, Pooyan; De Cecco, Carlo N.; Vliegenthart, Rozemarjin; Leonard, Tyler J.

Published in:
European Journal of Radiology

DOI:
10.1016/j.ejrad.2020.109428

Publication date:
2021

European Journal of Radiology 134 (2021) 109428

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: A validation study

Marly van Assen [a,g], Simon S. Martin [a,b], Akos Varga-Szemes [a], Saikiran Rapaka [c], Serkan Cimen [c], Puneet Sharma [c], Pooyan Sahbaee [d], Carlo N. De Cecco [g,e], Rozemarjin Vliegenthart [f,e], Tyler J. Leonard [a,e], Jeremy R. Burt [a,e], U. Joseph Schoepf [a,e,*]

[a] Division of Cardiovascular Imaging, Department of Radiology and Radiological Science, Medical University of South Carolina, USA
[b] Department of Diagnostic and Interventional Radiology, University Hospital Frankfurt, Frankfurt, Germany
[c] Medical Imaging Technologies, Siemens Healthineers, Princeton, NJ, USA
[d] Siemens Medical Solutions USA, Malvern, PA, USA
[e] Computed Tomography Research & Development, Siemens Healthineers, Forchheim, Germany
[f] University of Groningen, University Medical Center Groningen, Departments of Radiology, Groningen, the Netherlands
[g] Division of Cardiothoracic Imaging, Department of Radiology and Imaging Sciences, Emory University | Emory Healthcare, Inc., Atlanta, GA, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Cardiac
Chest
Computed tomography

ABSTRACT

Purpose: To evaluate deep-learning based calcium quantification on Chest CT scans compared with manual evaluation, and to enable interpretation in terms of the traditional Agatston score on dedicated Cardiac CT.
Methods: Automated calcium quantification was performed using a combination of deep-learning convolution neural networks with a ResNet-architecture for image features and a fully connected neural network for spatial coordinate features. Calcifications were identified automatically, after which the algorithm automatically excluded all non-coronary calcifications using coronary probability maps and aortic segmentation. The algorithm was first trained on cardiac-CTs and refined on non-triggered chest-CTs. This study used on 95 patients (cohort 1), who underwent both dedicated calcium scoring and chest-CT acquisitions using the Agatston score as reference standard and 168 patients (cohort 2) who underwent chest-CT only using qualitative expert assessment for external validation. Results from the deep-learning model were compared to Agatston-scores(cardiac-CTs) and manually determined calcium volumes(chest-CTs) and risk classifications.
Results: In cohort 1, the Agatston score and AI determined calcium volume shows high correlation with a correlation coefficient of 0.921($p < 0.001$) and R^2 of 0.91. According to the Agatston categories, a total of 67(70 %) were correctly classified with a sensitivity of 91 % and specificity of 92 % in detecting presence of coronary calcifications. Manual determined calcium volume on chest-CT showed excellent correlation with the AI volumes with a correlation coefficient of 0.923($p < 0.001$) and R^2 of 0.96, no significant difference was found ($p = 0.247$). According to qualitative risk classifications in cohort 2, 138(82 %) cases were correctly classified with a k-coefficient of 0.74, representing good agreement. All wrongly classified scans (30(18 %)) were attributed to an adjacent category.
Conclusion: Artificial intelligence based calcium quantification on chest-CTs shows good correlation compared to reference standards. Fully automating this process may reduce evaluation time and potentially optimize clinical calcium scoring without additional acquisitions.

Abbreviations: AI, artificial intelligence; CVDs, cardiovascular diseases; CCS, coronary artery calcium scoring.

* Corresponding author at: Department of Radiology and Radiological Science, Medical University of South Carolina, 25 Courtenay Drive, Charleston SC, USA.
E-mail address: schoepf@musc.edu (U.J. Schoepf).

https://doi.org/10.1016/j.ejrad.2020.109428
Received 18 September 2020; Received in revised form 6 November 2020; Accepted 18 November 2020
Available online 21 November 2020

1. Introduction

Cardiovascular diseases (CVDs) are a large contributor to the global mortality rate. A total of 17.9 million people die from CVDs every year, which accounts for 31 % of all global deaths [1]. Coronary artery calcium scoring (CCS) serves as a reliable tool and is generally recommended for use by several guidelines [2–4] for CVD risk assessment and to guide follow-up testing [5–8]. In general, the absence of coronary calcification in electrocardiography-triggered CT is associated with a very low cardiovascular risk and, thus, is commonly used to rule out coronary artery disease and will limit follow-up examinations. Increasing calcium scores, often divided into risk categories, reflect an increasing risk for cardiovascular events [9].

Conventional measurement of CCS requires the manual input of expert to identify coronary calcium lesions in each image section and is therefore labor intensive and time-consuming. Automated approaches can help reduce workload and reader variability, increasing the clinical applicability of CCS [10–12].

Standard CCS is performed on dedicated ECG-triggered non-contrast cardiac CT acquisitions. However, with the increased use of CCS and the increased numbers of CCS acquisitions, we also see an emerging role of non-contrast non-triggered chest CTs for the analysis of coronary calcium. The number of chest CT acquisitions for lung cancer screening are rapidly increasing with increasing evidence on the effectiveness of lung cancer screening programs [13,14]. Many risk factors are associated with CVD as well as with lung cancer, creating a large overlap in populations of interest. Even though these scans are performed without ECG gating, increasing the susceptibility for motion artifacts, they could potentially be used to simultaneously evaluate the individual risk of adverse cardiovascular events [15,16]. The use of these already clinically accepted acquisitions allows for risk assessment without the need for an additional acquisition, thereby reducing accumulated radiation dose.

Several studies have evaluated (semi)automatic methods for CCS in CT using standard cardiac calcium scoring CT [17–19] and chest CT [16, 20] acquisitions. In this study, a transfer learning approach is used to train the algorithm on both dedicated cardiac CCS acquisitions and on chest CT's, using an additional probability map to exclude non-coronary calcifications.

The aim of this study was to evaluate a novel deep learning-based algorithm for fully automated calcium scoring on non-contrast non-ECG-triggered chest CT compared to manual scoring and to permit interpretation of the AI-determined calcium volumes in terms of the well understood Agatston scoring on dedicated ECG-gated acquisitions.

2. Materials and methods

2.1. Patient population

This study performed validation of the software on 263 patients retrospectively included from a single-center cohort. A total of 95 patients underwent both dedicated CCS acquisitions and chest CT acquisition within 1.5 years were selected for cohort one. The median number of days between the cardiac and chest CT was 186 [76–383]. Only patients who had the calcium score noted in the cardiac CT report and with both chest CT and cardiac having diagnostic image quality were included. Patients who received cardiac intervention between the two scans and patients with metal assist devices such as pacemakers were excluded. This cohort was used to assess the quantitative properties of the AI algorithm (AI-Rad Companion Chest CT, Siemens Healthineers, Forchheim, Germany) by comparing the quantification of calcium volume on chest CT using the Agatston score from cardiac CT acquisitions as reference standard. A second cohort was formed by a total of 168patients, who underwent chest CT alone. This cohort consists out of consecutive patients who underwent chest CT imaging for non-cardiac purposes and was created to represent a clinically representable cohort for external AI validation of multiple algorithms. This cohort was used to assess the ability of the AI algorithm to mimic clinical qualitative coronary calcium analysis. This single-center, observational study was approved by the ethics committee of our university hospital and the need for informed consent was waived. Study was performed according to HIPAA regulations.

2.2. Scan protocol

2.2.1. Non-contrast ECG-Triggered cardiac CT

All data were acquired on a second or third-generation dual-source CT scanner (SOMATOM Force/Flash, Siemens Healthineers, Forchheim, Germany) or a Definition AS+ (Siemens Healthineers, Forchheim, Germany). CCS was performed via a prospectively ECG-triggered non-contrast sequential acquisition using the following parameters: tube voltage 120 kV, automated tube current modulation (CARE Dose4D, Siemens), reference tube current-time product of 80 mAs, collimation: $2 \times 128/192 \times 0.6$ mm, heart rate dependent pitch, and gantry rotation time 0.25−0.28 s. Most patients were scanned at 70 % ECG reconstruction percentage. Examinations were performed during inspiratory breath hold and in the craniocaudal direction. The scans were reconstructed with a routine filtered back projection (WFBP) algorithm, using a medium sharp convolution kernel, 3.0 mm section thickness, and an increment of 1.5 mm.

2.2.2. Non-contrast non-ECG-triggered chest CT

All data were acquired on a second or third-generation dual-source CT scanner (SOMATOM Flash/Force, Siemens Healthineers, Forchheim, Germany) or a Definition AS+ (Siemens Healthineers, Forchheim, Germany). Patients were scanned according to standard clinical protocol, representing a reflection of our clinical population, which would be ideal for clinical validation purposes. Here is an example of scanning parameters used for the Force CT scanner: tube voltage 100−120 kVp (based on BMI), automated tube current modulation (CARE Dose, Siemens), reference tube current-time product of 80 mAs, collimation: $2 \times 128/192 \times 0.6$ mm, gantry rotation time 0.25−0.28, slice thickness from 1 mm to 3 mm and slice spacing from 0.6 mm *to 3.0* mm."

2.3. Data analysis

2.3.1. Manual CACS

In cohort one, standard CCS, using the Agatston score, on non-contrast cardiac ECG-triggered CT acquisitions was performed manually using commercially available software (CT CaScoring, Siemens). Coronary artery calcifications were manually attributed to a coronary arteries as part of the clinical work-up and Agatston scores were reported in the clinical report by board-certified radiologist. Selection of coronary calcium and quantification of calcium volume was done manually on the non-triggered chest CT acquisitions. To evaluate the accuracy of risk assessment classification of the calcium quantification on chest CTs, patients were divided based on their Agatston score in standard risk categories (0, 1−10, 11−100, 101−400, or >400).

For the second cohort, scans were qualitatively analyzed by a cardiac radiologist with 19 years of experience (JRB). Each scan was given one of the following categories according to clinical guidelines: none, mild, moderate or severe calcifications. Our clinical standard reporting used these qualitative categories to assess coronary calcium. The categories (none, mild, moderate, or severe) were chosen to visually correlate with traditional CCS score groups [21]. The thresholds for assigning patients to a qualitative category were determined by the median calcium volume values of each category...

2.3.2. AI methodology

The coronary calcium volume was obtained using a deep-learning based algorithm (AI-Rad Companion Chest CT, Siemens Healthineers, Forchheim, Germany). The deep learning model has two components –

(i) a convolutional neural network, which takes the image patch and the local coronary territory map around each candidate voxel as inputs, and (ii) a dense neural network which operates on the coordinates of the voxel. The outputs of these two components are concatenated in the final layer of the network and a final prediction is made for each voxel to determine whether it belongs to the coronary arteries. Fig. 1 represents a schematic overview of the deep learning architecture. The output of the deep learning model is the total volume of the detected coronary calcifications. The gold standard for evaluation coronary calcifications is the Agatston scoring, which assigns different weights to each coronary calcification depending on the density of each calcification. Due to this weighting, there cannot be an exact correspondence between the calcium volume from the deep learning model to the Agatston scoring. However, understanding the relation between the calcium volume and the traditional Agatston risk scores on a representative clinical population could enable better interpretation of the model outputs in clinical routine.

The model is trained on a retrospective multi-center database of 1261 gated cardiac CT calcium scoring images to obtain a first model, which is then fine-tuned over a separate database of 500 non-gated, non-contrasted chest CT images. All the voxels in the cardiac region with intensity $>$ 130 HU are labeled by trained annotators and reviewed for accuracy by radiologists experienced in reading cardiac and chest CTs. The high intensity voxels in the coronary arteries corresponding to coronary calcium are labeled as positive samples, while the others (aortic/mitral calcification, any devices, high-intensity image noise etc.) are labeled as negative samples.

The entire training data set is first used to establish a coronary territory map in a heart coordinate system, to which each CT volume is mapped. This territory map serves to assign prior probabilities that a voxel belongs to the coronary arteries and is used as an additional input to the model.

For each candidate voxel, a small image patch is extracted centered around this voxel to represent the local spatial characteristics along with the location (x, y, z) of the voxel in the heart centric coordinate system. Voxels are only classified as calcifications if they are clustered resulting in a volume exceeding 5 mm [3], otherwise they are discarded as noise.

For qualitative risk classification the AI algorithm uses thresholds of 5, 250 and 1000 mm [3] of calcium volume to assign cases to mild, moderate and severe calcium volume categories.

2.4. Statistical analysis

In this study, the fully automated measurements on the non-triggered acquisitions were compared with 1) the manual measurements on the same acquisitions (cohort 1), and 2) the Agatston score of the corresponding ECG-triggered cardiac acquisitions (cohort 1) and 3) with manual qualitative calcium scoring using chest CT(cohort 2). Data was compared for the zero and non-zero groups and for different Agatston risk categories in patients who underwent both cardiac and chest CT. Independent and paired t-testing was used for data showing normal distribution, whereas the Wilcoxon signed-rank test was applied for data showing non-normal distribution. Pearson correlation coefficients were used to evaluate the correlation between the manually annotated CCS on both ECG-trigged and non-triggered CT acquisitions and the fully automated CCS method.

Statistical analyses were conducted using SPSS version 23 (IBM, Armonk, New York). A p-value $<$ 0.05 was considered statistically significant. Continuous variables are represented as mean (standard deviation [SD]) or median (interquartile range [IQR]), depending on their distribution (tested with Shapiro Wilkes test). Categorical data is displayed as absolute frequencies and proportions.

3. Results

3.1. Patient demographics

A total of 95 patients with both dedicated CCS and chest acquisitions were included for quantitative validation purposes.

A total of 168 patients undergoing chest CT only were included for the purpose of comparing qualitative calcium analysis. There were 66 (39 %), 53 (32 %), 25 (15 %) and 24 (14 %) patients with a no, mild, moderate and severe calcium classification according to expert opinion.

3.2. Agatston scores and AI determined chest CT calcium volumes

The median Agatson score of all 95 patients (cohort 1) undergoing cardiac and chest CT was 16.00 [IQR: 0.00−304.00]. Of these, 41 % (39/95) had an Agatston score of zero. The remaining percentage of patients fell in the following risk categories (1−10, 11−100, 101−400, or >400) 10 % (9/95), 13 % (12/95), 18 % (17/95), and 19 % (18/95), respectively.

According to the AI algorithm, the median calcium volume of all patients was 18.27 [IQR: 0.00−197.10]. An overview of all CACS for each risk category is given in Table 1. The Agatston score and AI determined calcium volume shows high correlation with a correlation coefficient of 0.921 ($p < 0.001$) and an R^2 of 0.91, see Fig. 2. According to the Agatston risk categories, a total of 67 (70 %) were correctly classified. Table 2 shows the confusion matrix for all calcium classifications based on Agatston score and AI determined calcium volume. It shows that only 3 cases (3 %) were falsely classified as 0 and in only 5 % the prediction was more than one category off.

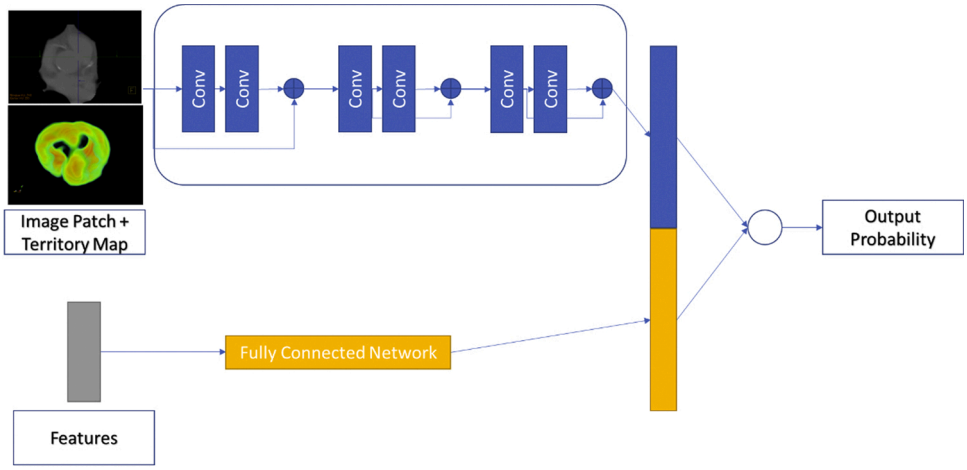

Fig. 1. Schematic overview of the AI architecture.

Table 1
Overview of Agatston scores and Calcium Volumes.

	Manual Cardiac CT (Agatston Score)	Manual Chest CT (Calcium volume)	AI Chest CT (Calcium volume)	p-values Manual vs. AI (Calcium volume)
n = 95				
Median (IQR)	16.00 [0.00–304.00]	13.20 [0.00–189.00]	18.27 [0.00–197.10]	0.247
Risk Categories:				
0 *(n = 39)*	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.144
1–10 *(n = 7)*	7.00 [3.50–10.00]	4.60[2.70–24.50]	0.00 [0.00–23.37]	0.866
11–100 *(n = 12)*	20.00 [16.25–23.87]	31.50 [6.40–59.05]	23.20 [6.76–38.32]	0.241
101–400 *(n = 17)*	185.00 [139.47–227.50]	132.40 [85.45–187.00]	119.85 [74.86–186.70]	0.177
400+ *(n = 20)*	907.50 [509.75–1372.50]	904.75[487.30–1285.60]	760.47 [352.73–1267.21]	0.526

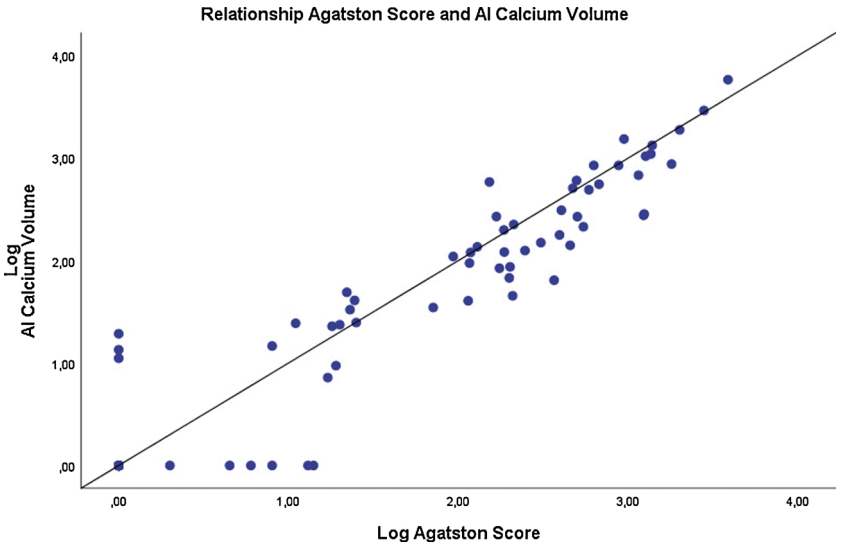

Fig. 2. Relationship between Agatston scores and AI determined calcium volume, showing excellent correlation 0.921 with an R2 of 0.91.

Table 2
Risk Category Agreement between Agatston (ref) and AI Calcium Volume.

Agatston\AI Volume	0	1-10	11-100	101-400	400+
0	36	1	2	0	0
1-10	3	0	3	1	0
11-100	2	0	7	1	0
101-400	0	0	7	9	1
400+	0	0	0	5	15

3.3. Volume vs. Volume

An overview of all calcium volumes (manually and AI determined) in cohort 1 patients for each separate risk category (based on the Agatston score) is given in Table 1. The manual determined, median calcium volume on chest CT was 13.20 [IQR: 0.00–189.00] and showed excellent correlation with the AI determined calcium volumes with a correlation coefficient f 0.923 ($p < 0.001$) and an R^2 of 0.96, see Fig. 3. Manual and AI determined calcium volumes were not significantly different ($p = 0.247$).

Bland Altman plot shows data well within the limits of agreement (Fig. 4) with an increasing error at higher calcium volumes.

3.4. Qualitative clinical assessment

In cohort 2, according to the expert reading 66 (40 %), 53 (32 %), 25 (15 %) and 24 (14 %) had no, mild, moderate or severe calcium volumes, respectively. According to the fully automated AI algorithm 67 (40 %), 56 (33 %), 20 (12 %) and 25(15 %) had no, mild, moderate or severe calcium volumes. In total 138 (82 %) cases were correctly classified with a kappa coefficient of 0.74 representing good agreement. All wrongly classified scans (30 (18 %)) were attributed to an adjacent category, see Table 3 for the full classification matrix. Comparing no calcium present vs calcium present, 6 (10 %) were considered mild where the expert scored them as no calcium present and 7 (11 %) cases were classified as no calcium present according to the AI algorithm where the expert stated that there was calcium present.

Fig. 5 shows an example of the AI segmentations used for calcium volume measurements, visualizing how the AI based aortic segmentation results in successful exclusion of aortic calcification for the determination of coronary calcium.

4. Discussion

A method for fully automated coronary calcium scoring from ECG-gated and non-gated chest CT data based on deep-learning models is presented and validated in this study. Two neural networks were combined; a convolution neural network with a ResNet-architecture for

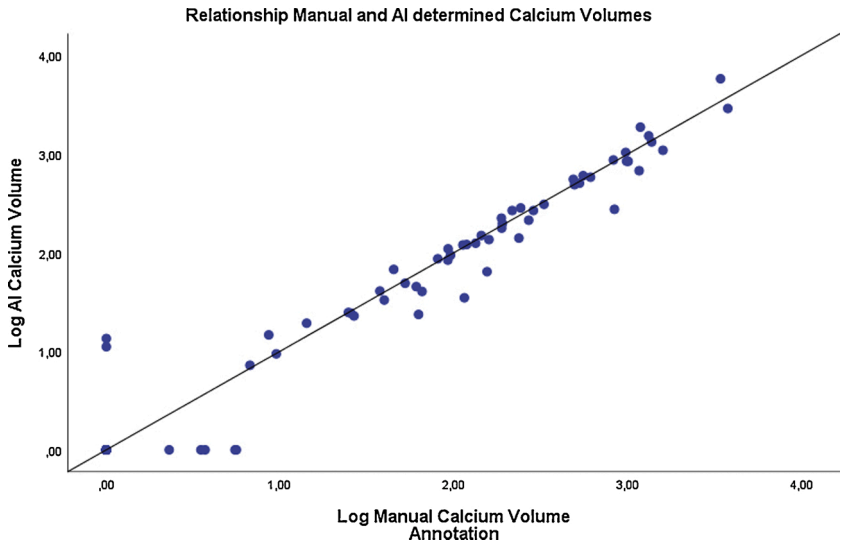

Fig. 3. Relationship between manual and AI determined calcium volume, showing excellent correlation 0.923 with an R2 of 0.96.

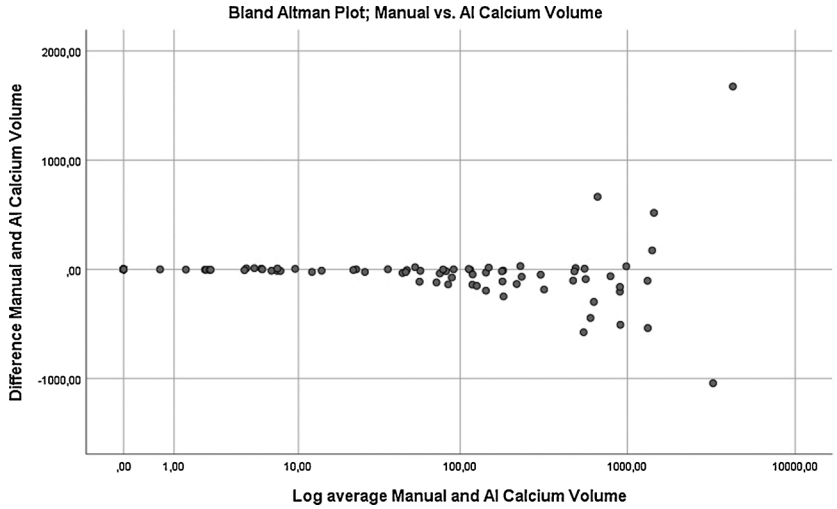

Fig. 4. Bland Altman plot, showing good agreement between manual and AI determined calcium volume from chest CTs. The absolute difference increases at higher absolute values. The x-axis show the log transformed average calcium volumes.

Table 3
Category agreement between manual qualitative assessment and AI determined calcium volume.

Expert\AI Volume	No	Mild	Moderate	Severe
No	**60**	6	0	0
Mild	7	**44**	0	0
Moderate	0	6	**14**	5
Severe	0	0	4	**20**

image features and a fully connected neural network for spatial coordinate features. A coronary probability map and deep-learning based aortic segmentation were used to exclude all non-coronary calcifications. Validation of this algorithm on ECG-gated cardiac CTs showed excellent correlation (0.921−0.923) between manual and AI determined calcium volumes and traditional Agatston scores. There were no significant differences between manual and AI determined calcium volumes. Using an AI algorithm on non-gated chest CTs, 70 % of patients were correctly assigned to the right Agatston risk category and 82 % were correctly classified according to qualitative calcium categorization used in clinical practice. When compared to Agatston scoring as ground truth for the presence/absence of coronary artery calcifications, the deep learning model achieved a sensitivity of 91 % (49/54 patients) and a specificity of 92 % (36/39 patients).

The application of CAC quantification on CT images has the potential to aid in the prediction of all-cause mortality and cardiovascular events and could help guide early phase therapy [22,23] Recent studies have shown that lung cancer screening, using chest CT, reduced lung cancer mortality and is likely to be implemented in a wide range of countries [13,14,24,25]. Aging and smoking are important risk factors for lung cancer and CVD, resulting in a large overlap of patients benefitting from CACS and lung cancer screening [26]. In the National Lung Screening

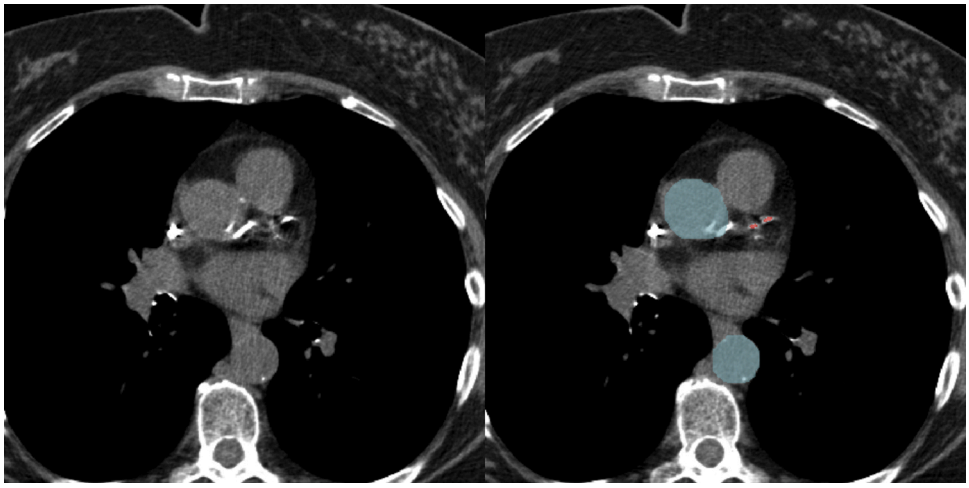

Fig. 5. Example of a chest CT image (left) and the same CT image with automated AI based aortic (blue) and coronary calcium (red) segmentation. Using an AI based aortic segmentation step ensures the exclusion of aortic calcium, as is demonstrated in this example.

Trial (NLST), CVD was the leading cause of mortality [27]. Automated quantification of CAC on chest CT acquisitions enables simultaneous evaluation of both lung cancer and cardiovascular risk, without additional costs and radiation exposure for patients and without increasing the workload at the Radiological department.

The proposed algorithm has been trained on both non-contrast ECG-triggered cardiac and non-ECG triggered chest CTs, making use of transfer learning to reduce the number of datasets needed while maintaining high accuracy. Chest CT's, most of which are non-ECG triggered, have higher noise levels and are more susceptible to motion artifact compared their ECG-triggered counterparts. For this reason, instead of determining Agatston scores, this algorithm focuses on the quantification of coronary calcium volume. Even though Agatston scoring might not be reliable in low-dose non-ECG-triggered scans, the quantification of coronary calcium volume enables estimation of cardiovascular risk [28]. It is expected that with increased motion on chest CTs the absolute volume is increased while the mass is expected to decrease. Using an Agatston score, including both parameters will be effected by both. In order to avoid overestimation, the thresholds of pixels involved for which a calcification is classified as positive has been adjusted in order to compensate for the increased motion and noise levels. Future research should investigate whether it is possible to convert these volumes to a traditional Agatston score.

Although our study shows an excellent, according to the quantitative analysis, comparing our AI quantification to the Agatston score risk categories, 30 % of cases was misclassified. The misclassification to the no calcium category will have the largest impact on patient treatment, since these patients will be considered to have no/little cardiac risk. Misclassifications of higher risk categories, although misrepresenting the risk, have little effect on the clinical follow-up and treatment of these patients However, the confusion matrix showed that only 3 cases (3 %) were falsely classified as 0 and in only 5 % the prediction was more than one category off. A recent meta-analysis showed that chest CT could not replace cardiac CT for screening purposes due to an average of 8.8 % false negatives on chest CT [29]. They also showed that a small percentage of subjects (up to 1.3 %) with zero CS in non-triggered CT had cardiovascular death or events. Their reported false negative percentage is higher than the percentage found in this study. According to a study by Xia et al. [30] on a large prospective cohort comparing Agatston scores on cardiac and chest CT's shows that 6.5 % of cases were misclassified, however, they used a very controlled population with no variety in CT scanners (only high-end) and imaging protocols in an asymptomatic screening cohort. In addition they show that misclassifications can increase to 13 % in cases of BMI $>$ 30. Our clinical cardiac populations, as used in this study with symptomatic patients undergoing cardiac CT imaging, are known for having an increased BMI. However, the population chosen for this study is a representation of clinical practice. Future studies should provide guidance whether chest CT based calcium quantification, whether or not used with AI should be used in high BMI patients. The AI algorithm could improve performance by performing an additional training on high BMI patients. In addition as mention before, adaptation of thresholds of pixels involved and minimum HU value of these pixel could further optimize calcium quantification.

In addition to our comparison with the Agatston score, our qualitative assessment also showed an 11 % misclassification to the zero calcium category. In the current study, the categorization of the AI based calcium quantification was performed using the medians of each category as assessed by the expert. However, optimization of these thresholds, in favor of correctly classifying the zero calcium category could be proven useful for screening purposes. A recent meta-analysis [29] has shown 8.8 % of false negative calcium score scans and concludes that chest should not be used for calcium screening according to the same guidelines as cardiac CT. The quantitative analysis of coronary calcium on chest CT however, still gives an indication of cardiovascular health in patients who receive chest imaging for non-cardiac purposes. It should be mentioned that in non-cardiac patients, a positive calcium score can be used as a tool to further assess cardiac health rather than clearing them of cardiovascular risk based on negative calcium score.

There are several studies performed on (semi) automated techniques for CAC quantification; however, there are only a few fully automated applications [11]. Wolterink et al. evaluated automated CCS methods for non-contrast CT and cardiac CTA [17,31], De Vos et al. proposed an algorithm on both cardiac and chest CT studies [32], while two other studies by Ebersberger et al. [33] and Ahmed et al. [34] investigated fully automated CAC analysis on contrast-enhanced CTA studies and standard non-contrast CT scans. Previous research on deep-learning based algorithms for automated detection of calcium on chest CTs show similar correlation to Agatston score or manual analysis with coefficients ranging from 0.93−0.98 [16,32,35–37] compared to 0.91−0.93 in the current study. Similar correlations are shown in studies on dedicated cardiac scans using ECG triggering with correlations ranging between 0.94−0.97. [17,31,38].

Our AI Algorithm uses a prior likelihood model for coronary territories to help with the calcium detection, similar as to the algorithm reported by Isgum et al. [20], however, the architecture of the underlying models is slightly different (Isgum et al. use a support vector classifier whereas our architecture uses convolutional neural networks). Another key difference of the current algorithm compared to previously

reported approaches is that, to the best of our knowledge, our model is the only one which is commercially available (AI Rad Companion Chest CT, Siemens Healthineers), making this the only study that performs a true external validation of an AI algorithm for calcium quantification on chest CT on a clinical representative population. True external validation is necessary to prove the workings of an AI algorithm in a clinical setting, using a representable population.

In addition to other AI approaches reported, the current approach also uses an additional model to reduce false positives calcium detection. One of the main reported sources of false positive calcium classifications is the classification of aortic, annular and valvular calcium [35]. Using a probability map and deep-learning based aortic segmentation to exclude non-coronary calcifications, the current algorithm reduces false positive classification of calcium. This is demonstrated by the relative underestimation of calcium volume by the AI algorithm. This underestimation of a fully automated approach determining calcium volumes compared to Agatston scores is also reported in a previous study by Takx et al. However, when compared to the inter-scan variability in non-ECG trigged chest CT acquisitions the errors of the automatic scoring are similar to those by manual expert scoring [28].

There are several limitations that deserve to be discussed. Although the current algorithm was trained and tested on multicenter data, the validation was performed on data originating from one institute. Future validation on a multi-center dataset should prove the generalizability of this algorithm. This study compared Agatson scores from dedicated cardiac scans with volumes from chest CTs. Although calcium volumes can be used for cardiovascular risk prediction, this parameter is not as standardized and thoroughly investigated as the Agatston score. Future research should be aimed in establishing a chest CT specific score, whether this is the Agatston score or not, for cardiovascular risk prediction.

In conclusion, deep-learning based automated calcium quantification on chest CT show excellent correlation with manual calcium volume quantification on chest CTs and with Agatston scored from cardiac CTs. Automated analysis could increase the workflow efficacy and help deal with the increasing number of acquisition requests in order to assist in the increased workforce of radiologists. This would be especially interesting in screening situations.

Disclosures

Dr. Schoepf receives institutional research support from Bayer, Bracco, Guerbet, and Siemens. Dr. Schoepf has received consulting fees and or speaker honoraria from Bayer, GE, Elucid BioImaging, Heartflow Inc., and Siemens. UMCG receives institutional research support from Siemens. Dr. Burt is owner of YellowDot Innovations, holds approved U. S. patents (USPTO 9,968,257 and 10,548,674), and receives consulting fees from Hyland Healthcare. The other authors have no conflict of interest to disclose.

Disclaimer

The products/features mentioned herein are not commercially available in all countries. Their future availability cannot be guaranteed.

Credit author statement

All authors contributed equally on the revisions of this manuscript and approved this final version being submitted.

Declaration of Competing Interest

Dr. Schoepf receives institutional research support from Bayer, Bracco, Guerbet, and Siemens. Dr. Schoepf has received consulting fees and or speaker honoraria from Bayer, GE, Elucid BioImaging, Heartflow Inc., and Siemens. UMCG receives institutional research support from Siemens. Dr. Burt is owner of YellowDot Innovations, holds approved U. S. patents (USPTO 9,968,257 and 10,548,674), and receives consulting fees from Hyland Healthcare. The other authors have no conflict of interest to disclose. Dr. De Cecco receives research funding from Siemens. Dr. Varga-Szemes receives research funding from Siemens and is consultant for Bayer and Elucid Bioimaging.

References

[1] WHO, World Health Organization, Key facts Cardiovascular Diseases (CVDs), 2017. Accessed March 18, 2019, https://www.who.int/cardiovascular_diseases/en/.
[2] D.C. Goff, D.M. Lloyd-Jones, G. Bennett, et al., Prevention guidelines tools CV risk calculator, Circulation 129 (25 suppl. 2) (2013) S49–S73, 2014; doi:1161/01.cir.0000437741.48606.98.
[3] S.D. Fihn, J.M. Gardin, J. Abrams, et al., Practice guideline 2012 ACCF / AHA / ACP / AATS / PCNA / SCAI / STS guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary, Circulation 126 (2012) 3097–3137, https://doi.org/10.1161/CIR.0b013e3182776f83.
[4] J. Knuuti, W. Wijns, I. Chairperson, et al., ESC Guidelines for the diagnosis and management of chronic coronary syndromes, Eur. Heart J. 2019 (2019) 1–71, https://doi.org/10.1093/eurheartj/ehz425.
[5] T.S. Polonsky, R.L. Mcclelland, N.W. Jorgensen, et al., Coronary artery calcium score and risk classification for coronary heart disease prediction, JAMA 303 (16) (2010) 1610–1616, https://doi.org/10.1001/jama.2010.461.
[6] D.H. O'Leary, M. Szklo, N.D. Wong, et al., Coronary calcium as a predictor of coronary events in four racial or ethnic groups, N. Engl. J. Med. 358 (13) (2008) 1336–1345, https://doi.org/10.1056/nejmoa072100.
[7] K. Nasir, M. Clouse, Role of nonenhanced multidetector Ct coronary artery calcium testing in asymptomatic and symptomatic individuals, Radiology 264 (3) (2012) 637–649, https://doi.org/10.1148/radiol.12110810.
[8] J. Yeboah, R.L. McClelland, T.S. Polonsky, et al., Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals, JAMA 308 (8) (2012) 788–795, https://doi.org/10.1001/jama.2012.9624.
[9] M. Oudkerk, A.E. Stillman, S.S. Halliburton, et al., Coronary artery calcium screening: current status and recommendations from the European Society of Cardiac Radiology and North American Society for Cardiovascular Imaging, Eur. Radiol. (2008).
[10] M. van Assen, I. Banerjee, C.N. De Cecco, Beyond the artificial intelligence hype: what lies behind the algorithms and what we can achieve, J. Thorac. Imaging 35 (Suppl. 1:S3–S10) (2020), https://doi.org/10.1097/RTI.0000000000000485.
[11] C.B. Monti, M. Codari, M. van Assen, C.N. De Cecco, R. Vliegenthart, Machine learning and deep neural networks applications in computed tomography for coronary artery disease and myocardial perfusion, J. Thorac. Imaging 35 (Suppl. 1: S58–S65) (2020), https://doi.org/10.1097/RTI.0000000000000490.
[12] A.M. Fischer, B. Yacoub, R.H. Savage, et al., Machine learning/deep neuronal network: routine application in chest computed tomography and workflow considerations, J. Thorac. Imaging 35 (May) (2020) S21–S27, https://doi.org/10.1097/RTI.0000000000000498.
[13] H.J. de Koning, C.M. van der Aalst, P.A. de Jong, E.T. Scholten KN, M. A. Heuvelmans, J.-W.J. Lammers, C. Weenink, U. Yousaf-Khan NH, S. van' t Westeinde, M. Prokop, W.P. Mali FAAMH, P.M.A. van Ooijen, J.G.J.V. Aerts, M.A. den Bakker ET, J. Verschakelen, R. Vliegenthart, J.E. Walter, Kten Haaf HJMG, M. Oudkerk, Reduced lung-cancer mortality with volume CT screening in a randomized trial, N. Engl. J. Med. (2020) 1–11, https://doi.org/10.1056/NEJMoa1911793.
[14] M. Oudkerk, A. Devaraj, R. Vliegenthart, et al., European position statement on lung cancer screening, Lancet Oncol. Elsevier Ltd 18 (12) (2017) e754–e766, https://doi.org/10.1016/S1470-2045(17)30861-6.
[15] I. Išgum, M. Prokop, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Automatic coronary calcium scoring in low-dose chest computed tomography, IEEE Trans. Med. Imaging 31 (12) (2012) 2322–2334, https://doi.org/10.1109/TMI.2012.2216889.
[16] R.A.P. Takx, P.A. De Jong, T. Leiner, et al., Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability, PLoS One 9 (3) (2014).
[17] J.M. Wolterink, T. Leiner, R.A.P. Takx, M.A. Viergever, I. Isgum, Automatic coronary calcium scoring in non-contrast-Enhanced ECG-triggered cardiac CT with ambiguity detection, IEEE Trans. Med. Imaging 34 (9) (2015) 1867–1878. United States.
[18] G. Brunner, D.R. Chittajallu, U. Kurkure, I.A. Kakadiaris, Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data, Int. J. Cardiovasc. Imaging 26 (7) (2010) 829–838. http://link.springer.com/10.1007/s10554-010-9608-1.
[19] A.M. Fischer, M. Eid, C.N. De Cecco, et al., Accuracy of an artificial intelligence deep learning algorithm implementing a recurrent neural network with long short-term memory for the automated detection of calcified plaques from coronary computed tomography angiography, J. Thorac. Imaging (2020).
[20] I. Isgum, M. Prokop, M. Niemeijer, M.A. Viergever, B. van Ginneken, Automatic coronary calcium scoring in low-dose chest computed tomography, IEEE Trans. Med. Imaging 31 (12) (2012) 2322–2334. United States.
[21] L. Azour, M.A. Kadoch, T.J. Ward, C.D. Eber, A.H. Jacobi, Estimation of cardiovascular risk on routine chest CT: ordinal coronary artery calcium scoring as an accurate predictor of Agatston score ranges, J. Cardiovasc. Comput. Tomogr 11 (1) (2017) 8–15, https://doi.org/10.1016/j.jcct.2016.10.001. Elsevier Ltd.

[22] T.S. Polonsky, R.L. McClelland, N.W. Jorgensen, et al., Coronary artery calcium score and risk classification for coronary heart disease prediction, JAMA 303 (16) (2010) 1610–1616, https://doi.org/10.1001/jama.2010.461.
[23] L.J. Shaw, P. Raggi, E. Schisterman, D.S. Berman, T.Q. Callister, Prognostic value of cardiac risk factors and coronary artery calcium screening for all-cause mortality, Radiology 228 (3) (2003) 826–833, https://doi.org/10.1148/radiol.2283021006.
[24] R. Meza, K. Ten Haaf, C.Y. Kong, et al., Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials, Cancer 120 (11) (2014) 17, https://doi.org/10.1002/cncr.28623.
[25] J.K. Field, A. Devaraj, S.W. Duffy, D.R. Baldwin, CT screening for lung cancer: is the evidence strong enough? Lung Cancer 91 (2016) 29–35, https://doi.org/10.1016/j.lungcan.2015.11.003.
[26] O.M. Mets, R. Vliegenthart, M.J. Gondrie, et al., Lung cancer screening CT-based prediction of cardiovascular events, JACC Cardiovasc. Imaging 6 (8) (2013) 899–907, https://doi.org/10.1016/j.jcmg.2013.02.008.
[27] T.R. Church, W.C. Black, D.R. Aberle, et al., Results of initial low-dose computed tomographic screening for lung cancer, N. Engl. J. Med. 23 (21) (2013) 368, https://doi.org/10.1056/NEJMoa1209120.
[28] P.C. Jacobs, M.J.A. Gondrie, Y. Van Der Graaf, 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) (2012) 505–511.
[29] X. Xie, Y. Zhao, G.H. De Bock, et al., Validation and prognosis of coronary artery calcium scoring in nontriggered thoracic computed tomography: systematic review and meta-analysis, Circ. Cardiovasc. Imaging 6 (4) (2013) 514–521.
[30] C. Xia, M. Vonder, G.J. Pelgrim, et al., High-pitch dual-source CT for coronary artery calcium scoring: a head-to-head comparison of non-triggered chest versus triggered cardiac acquisition: high-pitch chest versus cardiac CT for calcium scoring, J. Cardiovasc. Comput. Tomogr. (2020).
[31] J.M. Wolterink, T. Leiner, B.D. de Vos, R.W. van Hamersvelt, M.A. Viergever, I. Išgum, Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks, Med. Image Anal. 34 (2016) 123–136, https://doi.org/10.1016/j.media.2016.04.004. Elsevier B.V.
[32] B.D. de Vos, J.M. Wolterink, T. Leiner, P.A. de Jong, N. Lessmann, I. Isgum, Direct automatic coronary calcium scoring in cardiac and chest CT, IEEE Trans. Med. Imaging 38 (9) (2019) 2127–2138, https://doi.org/10.1109/TMI.2019.2899534.
[33] U. Ebersberger, D. Eilot, R. Goldenberg, et al., Fully automated derivation of coronary artery calcium scores and cardiovascular risk assessment from contrast medium-enhanced coronary CT angiography studies, Eur. Radiol. 23 (3) (2013) 650–657, https://doi.org/10.1007/s00330-012-2652–2656.
[34] W. Ahmed, M.A. de Graaf, A. Broersen, et al., Automatic detection and quantification of the Agatston coronary artery calcium score on contrast computed tomography angiography, Int. J. Cardiovasc. Imaging 31 (1) (2014) 151–161, https://doi.org/10.1007/s10554-014-0519–4.
[35] I. Išgum, M. Prokop, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Automatic coronary calcium scoring in low-dose chest computed tomography, IEEE Trans. Med. Imaging 31 (12) (2012) 2322–2334, https://doi.org/10.1109/TMI.2012.2216889.
[36] R. Shadmi, V. Mazo, O. Bregman-Amitai, E. Elnekave, Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT, Proc. – Int. Symp. Biomed. Imaging. (2018-April(Isbi)) (2018) 24–28, https://doi.org/10.1109/ISBI.2018.8363515.
[37] Carlos Cano-Espinosa, Germán González, George R. Washko, Miguel Cazorla, E. Raúl San José, Automated Agatston score computation in non-ECG gated CT scans using deep learning, Proc. SPIE. Int. Soc. Opt. Eng. 176 (1) (2017) 139–148, https://doi.org/10.1016/j.physbeh.2017.03.040.
[38] S.S. Martin, M. van Assen, S. Rapaka, et al., Evaluation of a deep learning–based automated CT coronary artery calcium scoring algorithm, JACC Cardiovasc. Imaging 13 (2 Pt 1) (2019), https://doi.org/10.1016/j.jcmg.2019.09.015.