


# Fully Automated Artery-Specific Calcium Scoring Based on Machine Learning in Low-Dose Computed Tomography Screening

## Vollautomatisches arterienspezifisches Kalzium-Scoring mittels maschinellem Lernen im Low-Dose CT-Screening

### Authors

Moritz T. Winkelmann<sup>1</sup>, Johann Jacoby<sup>2</sup>, Chris Schwemmer<sup>3</sup>, Sebastian Faby<sup>4</sup> , Patrick Krumm<sup>1</sup>, Christoph Artzner<sup>1</sup>, Malte N. Bongers<sup>1</sup> 

### Affiliations

- 1 Department for Diagnostic and Interventional Radiology, Eberhard Karls Universität Tübingen, Tübingen, Germany
- 2 Institute of Clinical Epidemiology and Applied Biometry, Eberhard Karls Universität Tübingen, Tübingen, Germany
- 3 Siemens Healthcare GmbH, Forchheim, Siemens Healthcare GmbH, Forchheim, Germany
- 4 Computed Tomography, Siemens Healthcare GmbH, Forchheim, Germany

### Key words

computed tomography, calcium scoring, coronary artery calcium, diagnostic accuracy, machine learning

received 01.10.2021

accepted 21.11.2021

published online 26.01.2022

### Bibliography

Fortschr Röntgenstr 2022; 194: 763–770

DOI 10.1055/a-1717-2703

ISSN 1438-9029

© 2022, Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

### Correspondence

Herr PD Dr. Malte N. Bongers

Abteilung für diagnostische und interventionelle Radiologie

Tübingen, Universitätsklinikum Tübingen,

Hoppe-Seyler-Straße 3, 72076 Tübingen, Germany

Tel.: +49/70 71/2 98 66 77

Fax: +49/70 71/29 46 38

malte.bongers@med.uni-tuebingen.de

### ABSTRACT

**Purpose** Evaluation of machine learning-based fully automated artery-specific coronary artery calcium (CAC) scoring software, using semi-automated software as a reference.

**Methods** A total of 505 patients underwent non-contrast-enhanced calcium scoring computed tomography (CSCT). Automated, machine learning-based software quantified the Agatston score (AS), volume score (VS), and mass score (MS) of each coronary artery [right coronary artery (RCA), left main

(LM), circumflex (CX) and left anterior descending (LAD)]. Identified CAC of readers who annotated the data with semi-automated software served as a reference standard. Statistics included comparisons of evaluation time, agreement of identified CAC, and comparisons of the AS, VS, and MS of the reference standard and the fully automated algorithm.

**Results** The machine learning-based software correlated strongly with the reference standard for the AS, VS, and MS (Spearman's rho > 0.969) ( $p < 0.001$ ), with excellent agreement (ICC > 0.919) ( $p < 0.001$ ). The mean assessment time of the reference standard was 59 seconds (IQR 39–140) and that of the automated algorithm was 5.9 seconds (IQR 3.9–16) ( $p < 0.001$ ). The Bland-Altman plots mean difference and 1.96 upper and lower limits of agreement for all arteries combined were: AS 0.996 (1.33 to 0.74), VS 0.995 (1.40 to 0.71), and MS 0.995 (1.35 to 0.74). The mean bias was minimal: 0.964–1.0429. Risk class assignment showed high accuracy for the AS in total (weighed  $\kappa = 0.99$ ) and for each individual artery ( $\kappa = 0.96$ – $0.99$ ) with corresponding correct risk group assignment in 497 of 505 patients (98.4%).

**Conclusion** The fully automated artery-specific coronary calcium scoring algorithm is a time-saving procedure and shows excellent correlation and agreement compared with the clinically established semi-automated approach.

### Key points:

- Very high correlation and agreement between fully automatic and semi-automatic calcium scoring software.
- Less time-consuming than conventional semi-automatic methods.
- Excellent tool for artery-specific calcium scoring in a clinical setting.

### Citation Format

- Winkelmann MT, Jacoby J, Schwemmer C et al. Fully Automated Artery-Specific Calcium Scoring Based on Machine Learning in Low-Dose Computed Tomography Screening. Fortschr Röntgenstr 2022; 194: 763–770

### ZUSAMMENFASSUNG

**Ziel** Evaluierung einer auf maschinellem Lernen basierenden vollautomatischen arterienspezifischen Software zur Bewer-

tion des Koronarkalkes (CAC), unter Verwendung einer halbautomatischen Software als Referenz.

**Methoden** Bei insgesamt 505 Patienten wurde eine nicht kontrastverstärkte Calcium-Scoring-Computertomografie (CSCT) durchgeführt. Eine automatisierte, auf Machine Learning basierende Software quantifizierte den Agatston-Score (AS), Volumen-Score (VS) und Massen-Score (MS) jeder Koronararterie (rechte Koronararterie [RCA], linke Koronararterie [LM], Ramus circumflexus [CX] und Ramus interventricularis anterior [LAD]). Ermittelte CAC der Reader, die die Daten mit einer halbautomatischen Software annotierten, dienten als Referenzstandard. Die Statistik umfasste Vergleiche der Auswertungszeit, Übereinstimmung der identifizierten CAC sowie Vergleiche von AS, VS und MS des Referenzstandards und vollautomatischen Algorithmus.

**Ergebnisse** Die auf maschinellem Lernen basierende Software korrelierte stark mit dem Referenzstandard für AS, VS und MS (Spearman's rho > 0,969) ( $p < 0,001$ ), mit hervorragender Übereinstimmung (ICC > 0,919) ( $p < 0,001$ ). Die mittlere Bewertungszeit des Referenzstandards betrug 59 s (IQR 39–140) und die des automatischen Algorithmus 5,9 s (IQR 3,9–16) ( $p < 0,001$ ). Die mittlere Differenz der Bland-Altman-Plots und die bei  $1,96 \times$  Standardabweichung definierten oberen und un-

teren Grenzen der Übereinstimmung für alle Arterien zusammen betragen: AS 0,996 (1,33 bis 0,74), VS 0,995 (1,40 bis 0,71), und MS 0,995 (1,35 bis 0,74). Der mittlere Bias war minimal: 0,964–1,0429. Die Risikoklassenzuordnung zeigte eine hohe Genauigkeit für den AS in Summe (gewichtetes  $\kappa = 0,99$ ) und für jede Arterie ( $\kappa = 0,96$ – $0,99$ ) mit entsprechender korrekter Risikogruppenzuordnung bei 497 von 505 Patienten (98,4%).

**Schlussfolgerung** Der vollautomatische arterienspezifische Koronarkalk-Scoring-Algorithmus ist ein zeitsparendes Verfahren und zeigt eine hervorragende Korrelation und Übereinstimmung mit dem klinisch etablierten halbautomatischen Ansatz.

#### Kernaussagen:

- Sehr hohe Korrelation und Übereinstimmung zwischen vollautomatischer und halbautomatischer Kalziumbewertungssoftware.
- Weniger zeitaufwendig als herkömmliche halbautomatische Verfahren.
- Hervorragendes Instrument zur arterienspezifischen Kalziumbestimmung im klinischen Alltag.

## Introduction

Coronary artery disease (CAD) is the leading cause of death worldwide [1–3]. Given the burden of CAD on patients and the health care system, early detection of the disease and prediction of the individual risk of developing cardiovascular events are crucial. Systematic research in this area has led to further developments in treatment and patient care and the possibility of individual risk assessment, which helps to optimize treatment and patient care [4, 5]. The current clinical guidelines in the US and Europe recommend calcium scoring computed tomography (CSCT) in selected asymptomatic individuals, typically at low to intermediate risk of CAD [6, 7].

Non-contrast-enhanced, ECG-triggered CSCT is performed at a low radiation dose and can determine the cardiovascular risk for each patient, using the well-established metrics Agatston score (AS), volume score (VS), and mass score (MS) [8]. The AS calculates calcium burden by multiplying the area of the lesion above a 130 HU threshold and VS is defined as the total number of voxels exceeding the threshold of 130 HU for the respective calcium region [8, 9]. Whereas VS and AS are intended as indirect indicators of coronary artery calcification (CAC), MS provides an actual quantitative measure and assesses the true mass of CAC [8].

Typically, radiologists use semi-automated software for evaluation, including manual detection and marking of coronary artery calcifications [10], supported by threshold-based, automated region-growing algorithms. Up to now, measurement of CAC requires manual input by a human operator to identify and assign calcified coronary lesions to the left main artery (LM), left anterior

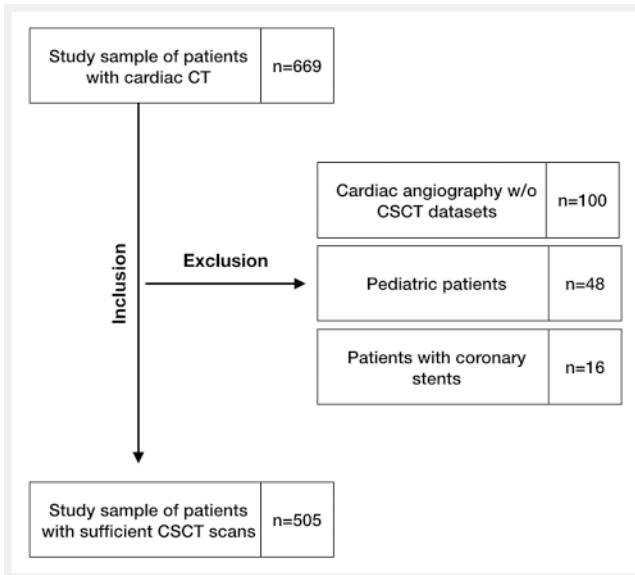
descending artery (LAD), circumflex artery (CX), or right coronary (RCA) artery [11, 12].

Due to the worldwide use of CSCT, there is a need to further improve and automatize the examination and post-processing workflow [13]. In recent years, developments in machine learning have led to improvements in automated systems for CSCT [10, 13–15]. With regard to determining the total calcium load of all coronary arteries, some studies have already shown promising results [10, 14]. Data regarding the performance of machine learning-based algorithms for the detection of CAC with identification of the particular coronary artery are limited. However, knowledge of the calcium load of each individual coronary vessel could have an impact on cardiovascular risk management. In fact, CAC of the LM and LAD was associated with increased mortality risk and CAC of the right coronary artery with decreased mortality risk [16, 17].

The aim of this retrospective single-center study was to evaluate novel machine learning-based software for fully automated calcium scoring with identification and evaluation of each coronary artery in non-contrast cardiac CT, as compared to a semi-automated post-processing tool serving as the standard of reference.

## Materials & Method

The local institutional review board approved this retrospective analysis of patient data. In this retrospective single-center study, patients and their baseline characteristics were retrospectively collected from the institutional database. A total of 505 patients with CSCT performed on a state-of-the-art CT scanner (SOMATOM Definition Flash or SOMATOM Force; Siemens Healthineers, Erlangen,



► **Fig. 1** STARD flowchart of patient inclusion.

► **Abb. 1** STARD-Diagramm der Studienpopulation.

Germany) between January 2013 and July 2020 were included. The exclusion criteria were cardiac CT without non-contrast-enhanced ECG-triggered calcium scoring datasets, pediatric cardiac CT datasets, and patients with intracoronary stents (► **Fig. 1**).

### Imaging protocol

All CSCT scans were performed on a state-of-the-art multidetector CT scanner (SOMATOM Definition Flash or SOMATOM Force; Siemens Healthineers, Erlangen, Germany). All images were acquired with automatic tube current modulation (CARE Dose 4D), automatic kV modulation (CARE kV), and a reference mAs of 60 and a reference kV of 120. For SOMATOM Force, the gantry rotation time was 0.25 s, the pitch was 3.2, and the collimation was  $0.6 \times 192$  mm. Reconstructions were computed with Sa36 kernel, a slice thickness of 3.0 mm, and an increment 1.5 mm. For SOMATOM Definition Flash, the gantry rotation time was 0.28 s, the pitch was 3.4, and the collimation was  $0.6 \times 128$  mm. Reconstructions were made with B35 f. kernel, a slice thickness of 3.0 mm, and an increment of 1.5 mm. If the patient's heart rate was above 65 bpm, a beta-blocker (5 mg Metoprolol, Recordati Pharma GmbH, Germany) was administered intravenously. Following CSCT, contrast-enhanced angio/cardiac CT was performed.

### Machine learning-based Calcium scoring software

The automated software was trained on 1261 anonymized datasets from routine coronary artery calcification examinations from multiple vendors, scanners, and from different hospitals. No training data sets were analyzed in the current study.

First, the standard 130 HU threshold is applied to the image to identify voxels as calcium candidates. For each candidate voxel, a small piece of image information as well as the voxel position in a cardiac coordinate system and some local features (e. g., HU value

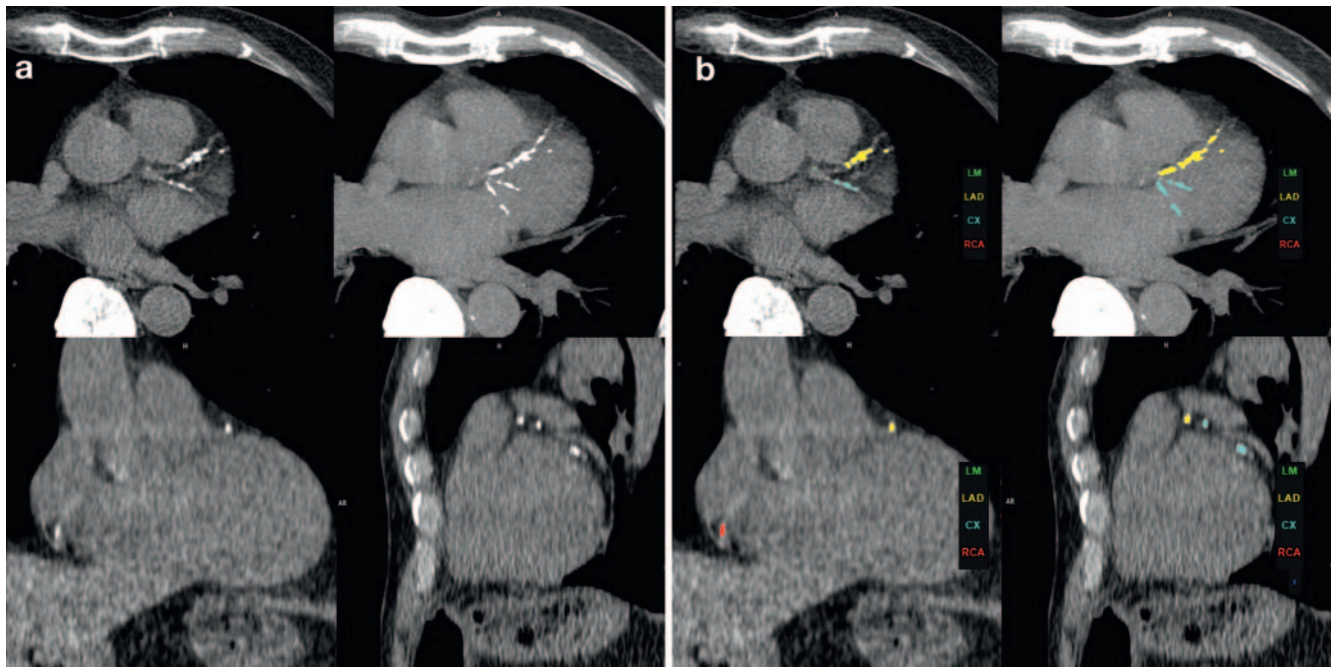
of the voxel) are extracted. The Deep Learning model works with two components, a convolutional neural network (CNN) with ResNet architecture that processes the image piece around the voxel and a dense network that processes the position in the heart coordinate system and the local features. The results of both networks are merged and plugged into a classifier that outputs the probability of coronary calcium for each voxel. If the average probability of a connected cluster is higher than a predefined threshold, it is marked in the application. The CNN is accompanied by an atlas trained with segmented coronaries from CTAs. Therefore, this component indicates whether a voxel is likely to be coronary or not, thus excluding heart valves, etc.

The next step is a deep learning algorithm that provides the position of the LM-LAD-LCX bifurcation. For this, the final classification of the branch is performed using a simple, fully connected neural network whose features include the spatial coordinates of each voxel identified as belonging to the coronary arteries and the coordinates of the voxel as a function of coronary bifurcation. This model yields 4 outcomes, namely the probability that the voxel belongs to the LM, LAD, CX, or RCA. The final assignment is made using a softmax function to determine the most likely position for each voxel.

### Calcium scoring and evaluation of the machine learning-based software

Semi-automated, clinically established post-processing software (syngo.via, version VB50 Siemens Healthineers) was utilized to generate the reference standard. All 505 CSCT scans were double-read by two radiologists in multiple sessions (Reader 1 with nine years and Reader 2 with four years of experience in cardiac CT diagnostics), and all differences in image interpretation were resolved by consensus. To avoid bias, both readers were blinded to the results of the automated software. As previously described in the literature [4, 10], to detect CAC, a threshold of  $> 130$  HU was determined on an area of  $\geq 1$  mm<sup>2</sup>, which corresponded to the default setting of the software. Calcified lesions of interest were manually identified and assigned to their respective coronary artery type (LM, LAD, CX, RCA). Regions were labeled to obtain the number of calcified lesions, the artery-based AS, VS, MS, and total AS, VS, MS. After loading the images into the software, the time measurement of the automatic system was started after the onset of the automatic assessment and stopped after the software displayed a score. The evaluation time of the reference standard included the location of all CACs and the correlation of the automatically derived number of CACs. The time that was required for the first reading was registered.

The individual scans were assigned to risk groups, which are standardized [18] and based on the AS. CAC 0: very low risk; CAC 1–10: low risk, CAC 11–100: moderate risk, CAC 101–400: moderately high risk, CAC  $> 400$ : high risk. The automated software was used on a regular daily routine diagnostic workstation (syngo.via, version VB50 Siemens Healthineers). All CSCT scans (n = 505) were analyzed with the machine learning-based automated software. The number of calcified lesions was registered and additionally assigned to the respective coronary artery. The AS, VS, and MS for the respective coronary artery and the total AS,



► **Fig. 2 a–b** Reconstructions in axial, axial thin-section MIP, coronary, and sagittal planes with calcifications in the LM, LAD, CX, and RCA. **a** Visually visible coronary calcifications before application of automatic calcium scoring software. **b** Calcium regions detected by the automated software and color-coded for the corresponding artery.

► **Abb. 2 a–b** Rekonstruktionen in der axialen Ebene, axialen Dünnschnitt-MIP, Koronar- und Sagittalebene mit Verkalkungen in der LM, LAD, CX und RCA. **a** Visuell sichtbare Koronarkalkablagerungen vor Anwendung der automatischen Kalzium-Scoring-Software. **b** Von der automatischen Software ermittelte und für die entsprechende Arterie farbkodierte Kalziumregionen.

VS, and MS were determined. The duration of the system run time was recorded. Subsequently, a double-check of the results was performed in which the number and location of the calcified lesions were reviewed. The only human interaction that was needed was for loading the images into the software (► **Fig. 2**).

**Statistics**

The available data were analyzed using SPSS (SPSS Statistics 26, IBM Corp., Armonk, New York, USA), R version 4.0.3 (The R Foundation for Statistical Computing, Vienna, Austria), in particular using the package Blandr [19]. Continuous variables are presented as mean ± standard or as the median and interquartile range (IQR) if non-normally distributed. The correlation and agreement between the standard reference and the machine learning-based software for coronary artery-based and total AS, VS, MS, and the number of lesions were calculated with Spearman’s rank correlation coefficient (ρ) and intraclass correlation coefficient (ICC). The reference standard and the machine learning-based automatic software were compared by way of a Bland-Altman procedure. The agreement was examined after recoding values of 0 to 0.06 and subsequent log transformation because of the right skewness of the data. Differences in risk classifications were assessed by weighted kappa analysis (κ). The time difference was determined using the Wilcoxon signed-rank test.

► **Table 1** Patient characteristics.

► **Tab. 1** Patientencharakteristika.

Variables	N (%) / mean ± SD
Patients	505
Women	132
Men	373
Age (years)	57.6 ± 12.6
BMI	25.4 ± 5.4

**Results**

A total of 505 patients were successfully included in the study based on the inclusion criteria: 132 (26.1%) women and 373 men (73.9%). The mean age was 57.6 ± 12.6 years (► **Table 1**).

The median time for the semi-automatic collection of data for the reference standard was 59 seconds (IQR, 39–140 sec) compared to the time of 5.9 seconds (IQR, 3.9–16 sec) required by the automatic machine learning-based algorithm (p < 0.001).

The correlation and agreement of the automatic algorithm and the reference standard concerning the number of calcified lesions were calculated by Spearman’s rank correlation coefficient and ICC

for the respective arteries (Spearman's rho >0.965; ICC >0.870) (p<0.001) (► **Table 2**).

The coronary artery calcium scoring results of the machine learning-based software correlated highly with the reference standard for the AS, VS, and MS for all four coronary arteries (Spearman's rho >0.969) (p<0.001). The Spearman's rho of the individual arteries can be found in ► **Table 2**.

The agreement of the machine learning-based software with the reference standard was evaluated using ICC. In terms of the AS, VS, and MS, the ICC was 0.983, 0.978, and 0.981, respectively, for the LM, 0.954, 0.953, and 0.957 for the LAD, 0.919, 0.922, and 0.924 for the CX, and 0.989, 0.989 and 0.989 for the RCA. The ICC for the total values of the AS, VS, and MS was 0.996, 0.995, and 0.992, respectively (p<0.001) (► **Table 2**).

The Bland-Altman plots mean difference (log-transformed, theoretical line of no bias y = 1) and 1.96 upper and lower limits of agreement for all arteries combined was: AS 0.996 (1.33 to 0.74), VS 0.995 (1.40 to 0.71), and MS 0.995 (1.35 to 0.74). The mean bias was minimal for the respective coronary arteries (0.964–1.0429). The values for the individual arteries are shown in ► **Fig. 3** and ► **Table 3**.

Weighted kappa analysis for risk class assignment showed high accuracy for the AS in total (weighted κ = 0.99) and for each artery (κ = 0.96–0.99). There were a total of 88 misclassifications with consecutive change of the total Agatston score. Most scans were incorrect within the low-risk category (CAC 1–10: n = 58) and moderate-risk category (CAC 11–100: n = 22). These minor errors had no effect on the assignment of the risk group and occurred mainly due to misregistration of image noise in the heart and adjacent structures. The fully automated software classified 497 of 505 patients (98.4%) into the correct risk category.

In five patients (1%) with significant errors in the moderate high-risk category (CAC 101–400), the software did not include calcification at the right coronary ostium (n = 1) or malfunctioned in differentiating between coronary and pericardial calcifications (n = 4), thus underestimating the calcium load. Significant overestimation of calcium load was observed in three patients in the high-risk group (CAC >400) due to erroneous inclusion of calcifications at the aortic root (n = 1), pericardium (n = 1), and mitral valve (n = 1) (► **Fig. 4a–c**).

## Discussion

In this study, the performance of novel machine learning-based fully automated post-processing software was evaluated for artery-based calcium scoring in cardiac CT, compared with clinically established semi-automated post-processing software serving as the standard of reference. Correlation, agreement, and risk classification were excellent for each artery and in total. Compared with the semi-automated approach, the fully automated analysis allows a tailored survey of each patient's calcium load to be collected in significantly less time.

For the coronary arteries separately and as a total, the correlation and agreement of the number of lesions, the AS, the VS, and the MS of the machine learning-based software were excellent compared with the reference standard. The Bland-Altman plot

► **Table 2** Measures of association between automatic algorithm and reference standard.

► **Tab. 2** Grad der Übereinstimmung zwischen automatischem Algorithmus und Referenzstandard.

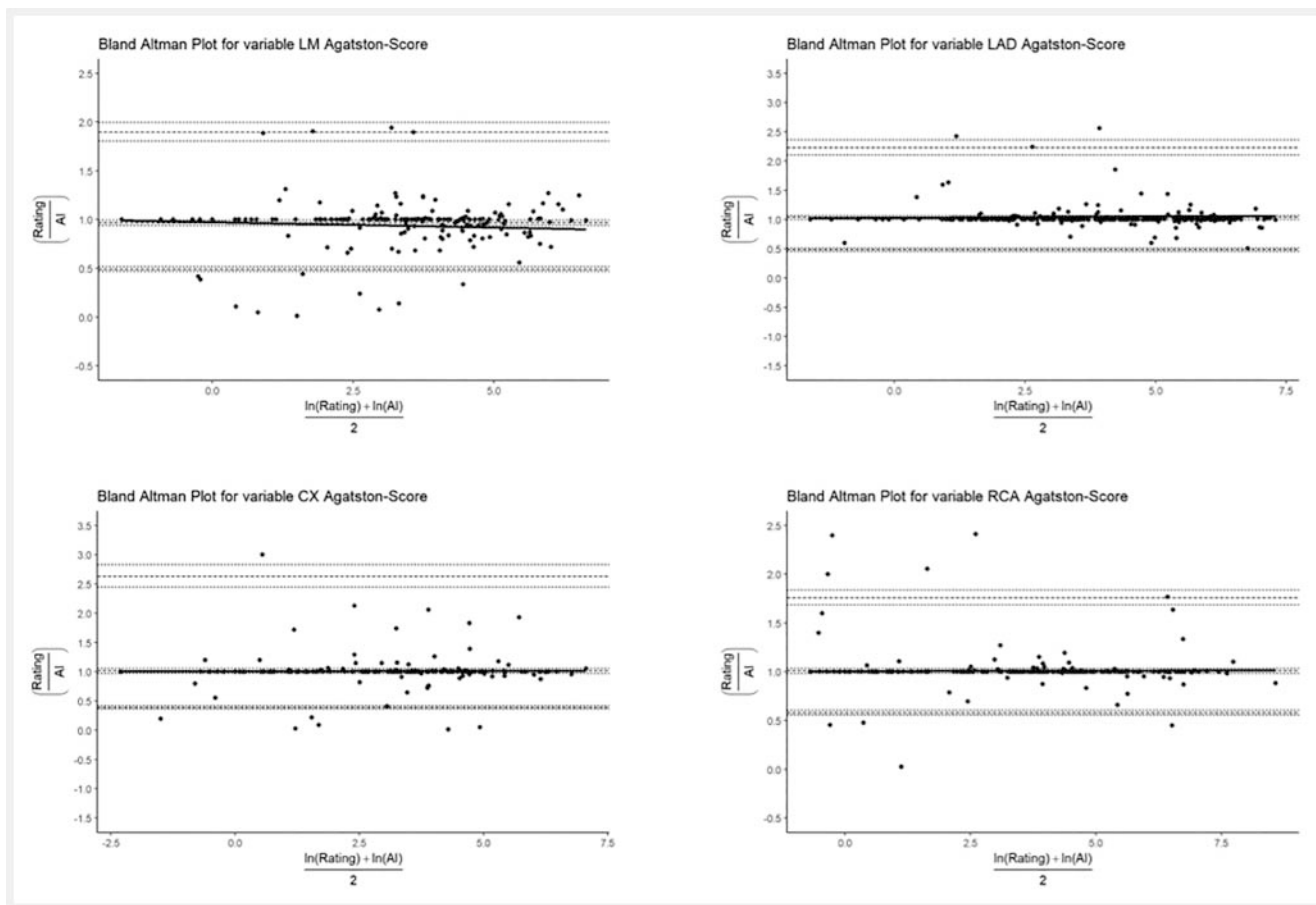
Measure	Spearman *	ICC*	95% CI ICC
LM number of lesions	0.965	0.870	[0.847–0.890]
LM volume (mm <sup>2</sup> )	0.982	0.978	[0.973–0.981]
LM equiv. mass (mg)	0.982	0.981	[0.977–0.984]
LM Agatston-score	0.982	0.983	[0.979–0.985]
LAD number of lesions	0.987	0.948	[0.938–0.956]
LAD volume (mm <sup>2</sup> )	0.996	0.953	[0.944–0.960]
LAD equiv. mass (mg)	0.996	0.957	[0.945–0.964]
LAD Agatston-score	0.996	0.954	[0.950–0.961]
CX number of lesions	0.966	0.952	[0.943–0.960]
CX volume (mm <sup>2</sup> )	0.969	0.922	[0.910–0.936]
CX equiv. Mmass (mg)	0.969	0.924	[0.910–0.936]
CX Agatston-score	0.970	0.920	[0.905–0.932]
RCA number of lesions	0.980	0.972	[0.967–0.977]
RCA volume (mm <sup>2</sup> )	0.986	0.990	[0.988–0.991]
RCA equiv. mass (mg)	0.986	0.990	[0.987–0.991]
RCA Agatston-score	0.986	0.990	[0.998–0.992]
Total number of lesions	0.995	0.977	[0.973–0.981]
Total volume (mm <sup>2</sup> )	0.999	0.995	[0.994–0.996]
Total equiv. mass (mg)	0.998	0.992	[0.991–0.993]
Total Agatston-score	0.999	0.996	[0.995–0.996]

LM = left main artery, LAD = left anterior descending artery, CX = circumflex artery, RCA = right coronary artery.

\* all p<0.001.

for the AS, VS, and MS showed a high level of agreement for all arteries. The Bland-Altman evaluation that was performed is based on the logarithmized values of the two measurements (automatic software and reference standard). This transformation is appropriate in the case of values that are highly right-skewed-distributed and downward-bounded. In our study, skewness of the data set was present, as 213 of 505 patients (42%) had a total AS of 0. Weighted kappa analysis provided accurate risk group categorization.

Several studies have already evaluated automated software for CSCT with comparable results regarding correlation and agreement for calcium scoring and risk category classification [10, 11, 20]. Due to differences in study design, data distribution, and quantitative assessment, comparisons are difficult. The larger number of patients in our study confirms the robustness of the



▶ **Fig. 3** Bland-Altman plots (log-transformed with back transformation) for LM, LAD, CX, and RCA. Mean of log (rating) and log (artificial intelligence) on the x-axis, Rating by humans/AI result ratio on the y-axis. The theoretical line of no bias is at  $y = 1$ . Dashed lines indicate bias and LOAs, and dotted lines indicate 95 % confidence bands. The solid line represents proportional bias. Observations with rating/AI ratios higher than the maximum value on the y-axis are omitted for presentation, while analysis used all available cases.

▶ **Abb. 3** Bland-Altman-Diagramme (log-transformiert mit Rücktransformation) für LM, LAD, CX und RCA. Mittelwert von log (Rating) und log (künstliche Intelligenz) auf der x-Achse, Verhältnis zwischen menschlichem Rating und AI-Ergebnis auf der y-Achse. Die theoretische Linie ohne Bias liegt bei  $y = 1$ . Die gestrichelten Linien zeigen die Verzerrungen und die LOAs an, die gepunkteten Linien die 95 %-Konfidenzbänder. Die durchgezogene Linie stellt den proportionalen Bias dar. Beobachtungen mit Rating/AI-Verhältnissen, die über dem Maximalwert auf der y-Achse liegen, wurden für die Darstellung ausgelassen, während für die Analyse alle verfügbaren Fälle verwendet wurden.

automated software for the evaluation of CSCT. In contrast to previous studies [10], exclusion of patients with metallic foreign bodies such as heart valve replacements and cardiac pacemakers was not necessary. The software’s CNN is trained to differentiate whether a voxel belongs to a coronary artery or metal implant.

The number of studies evaluating automated CSCT software with calcium load assignment for each coronary artery is limited [20]. Since the risk from calcium burden can vary for each coronary vessel, the excellent performance of artery-specific automated calcium score evaluation can contribute to time-efficient, cost-effective, tailored CAD screening [21]. The results of our study suggest that artery-specific automated calcium assessment software could be integrated into routine clinical practice for the quantification of coronary calcium with additional branch labeling. Since the software will be commercially available, widespread clinical implementation and workflow integration are anticipated and will hopefully yield the same results as our study.

We are aware that our study has limitations, mainly due to its retrospective nature, and we made every effort to create a strong reference standard with two independent, experienced readers. All CSCT scans were performed in a single center on two different CT scanners from the same vendor. It was already presumed that calcium scoring from other vendors might vary [22]. The automatic software was compared with semi-automatic software from the same vendor. However, the results of the semi-automatic software can be reproduced on other platforms [23]. Although this is one of the most extensive known studies evaluating automatic CAC scoring from CSCT scans, an even larger data set would undoubtedly lead to even more robust results.

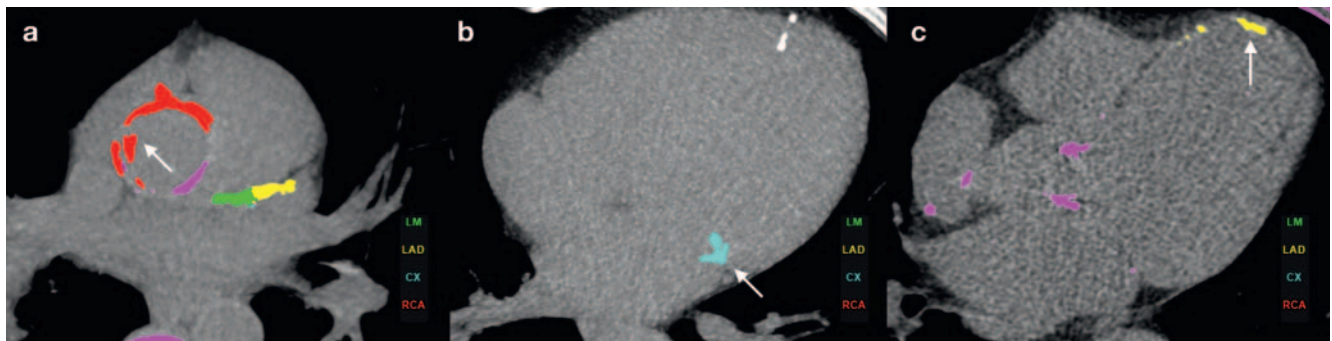
Despite the overall excellent performance of the algorithm, there were some outliers. Misclassification by the automated software occurred in five patients in the intermediate to high-risk group, with calcifications at the ostium of the right coronary artery not detected in one patient and partial failure to distinguish

► **Table 3** Bland Altman procedure with log-transformed measurement values, results in back-transformed (exponentiated).

► **Tab. 3** Bland-Altman-Methode mit log-transformierten Messwerten und rücktransformierten Ergebnissen (potenziert).

Measure	Mean bias	Upper limit of agreement	Lower limit of agreement	p for proportional bias
LM volume (mm <sup>2</sup> )	0.944	2.622	0.340	0.252
LM equiv. mass (mg)	0.961	2.006	0.460	0.117
LM Agatston-score	0.945	2.580	0.346	0.254
LAD volume (mm <sup>2</sup> )	1.041	2.182	0.496	0.422
LAD equiv. mass (mg)	1.047	2.630	0.417	0.771
LAD Agatston-score	1.044	2.640	0.481	0.344
CX volume (mm <sup>2</sup> )	1.047	3.848	0.285	0.745
CX equiv. mass (mg)	1.030	2.802	0.379	0.999
CX Agatston-score	1.050	3.841	0.287	0.972
RCA volume (mm <sup>2</sup> )	1.014	2.693	0.381	0.929
RCA equiv. mass (mg)	1.011	1.888	0.541	0.852
RCA Agatston-score	1.014	2.434	0.422	0.851
Total volume (mm <sup>2</sup> )	0.995	1.403	0.706	0.699
Total equiv. mass (mg)	0.995	1.347	0.735	0.485
Total Agatston-score	0.996	1.332	0.744	0.812

LM = left main artery, LAD = left anterior descending artery, CX = circumflex artery, RCA = right coronary artery.



► **Fig. 4 a–c** Reconstruction in axial planes after application of the automatic calcium scoring software in three different patients. Depicted is an overestimation of calcium load (arrows) by the automatic algorithm due to incorrect inclusion of calcifications at the aortic root **a**, mitral valve **b**, and in the pericardium **c**.

► **Abb. 4 a–c** Rekonstruktion in axialen Ebenen nach Anwendung der automatischen Kalzium-Scoring-Software bei 3 verschiedenen Patienten. Dargestellt ist eine Überschätzung der Kalziumbelastung (Pfeile) durch den automatischen Algorithmus durch fehlerhafte Einbeziehung von Verkalkungen an der Aortenwurzel **a**, der Mitralklappe **b** und im Perikard **c**.

between coronary calcification and calcification in the pericardium in the remaining patients. In a total of three patients, there was a misclassification into the high-risk group due to an overestimation of the calcium burden because of an incorrect detection of calcifications at the aortic arch, the pericardium, and the mitral valve. However, these distinct errors are not difficult to detect when reviewing the results and may therefore be of limited clinical relevance. For this reason, the results of the automated algo-

gorithm should always be verified by a human observer when used in routine clinical practice.

Furthermore, it would be beneficial to further develop the software to apply to non-ECG-triggered, standard CT thorax examinations. A number of studies have already addressed epidemiologic stratifications of coronary calcification on conventional chest CT [24–26]. However, the present study was designed to automatic-

ally assess coronary calcification on cardiac CT in a large population in a detail-oriented manner.

In conclusion, this study presented the validation of fully automated software for artery-specific detection of coronary calcification. The results showed excellent correlation and agreement between the automatic and the reference standard for three CAC scores and the number of coronary lesions in each coronary artery.

### CLINICAL RELEVANCE

- Coronary calcium load is known to predict cardiovascular risk, and its automatic and time-efficient determination is of clinical importance.
- The utilization of machine learning-based applications in clinical practice can improve workflow efficiency for frequent CT examinations, such as non-contrast-enhanced calcium scoring computed tomography.

### Conflict of Interest

S.F. and C.S. are employees of Siemens. All other authors declare that they have no conflict of interest.

### References

- [1] Nowbar AN, Gitto M, Howard JP et al. Mortality From Ischemic Heart Disease. *Circ Cardiovasc Qual Outcomes* 2019; 12: e005375
- [2] Hecht HS. Coronary artery calcium scanning: past, present, and future. *JACC Cardiovascular imaging* 2015; 8: 579–596
- [3] Lee H, Martin S, Burt JR et al. Machine Learning and Coronary Artery Calcium Scoring. *Curr Cardiol Rep* 2020; 22: 90
- [4] Agatston AS, Janowitz WR, Hildner FJ et al. Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American College of Cardiology* 1990; 15: 827–832
- [5] Divakaran S, Cheezum MK, Hulten EA et al. Use of cardiac CT and calcium scoring for detecting coronary plaque: implications on prognosis and patient management. *The British journal of radiology* 2015; 88: 20140594
- [6] Piepoli MF, Hoes AW, Agewall S et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Atherosclerosis* 2016; 252: 207–274
- [7] Greenland P, Alpert J, Beller G et al. American Heart Association Task Force on Practice Guidelines. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2010; 122: 2748–2764
- [8] Alluri K, Joshi PH, Henry TS et al. Scoring of coronary artery calcium scans: history, assumptions, current limitations, and future directions. *Atherosclerosis* 2015; 239: 109–117
- [9] Rumberger JA, Kaufman L. A Rosetta Stone for Coronary Calcium Risk Stratification: Agatston, Volume, and Mass Scores in 11,490 Individuals. *American Journal of Roentgenology* 2003; 181: 743–748
- [10] Sandstedt M, Henriksson L, Janzon M et al. Evaluation of an AI-based, automatic coronary artery calcium scoring software. *European radiology* 2020; 30: 1671–1678
- [11] Wolterink JM, Leiner T, De Vos BD et al. An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework. *Medical physics* 2016; 43: 2361–2373
- [12] Brunner G, Chittajallu DR, Kurkure U et al. Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data. *The international journal of cardiovascular imaging* 2010; 26: 829–838
- [13] Sharma P, Suehling M, Flohr T et al. Artificial Intelligence in Diagnostic Imaging: Status Quo, Challenges, and Future Opportunities. *J Thorac Imaging* 2020; 35: S11–S16
- [14] Martin SS, Assen Mv, Rapaka S et al. Evaluation of a Deep Learning – 2013; Based Automated CT Coronary Artery Calcium Scoring Algorithm. *JACC: Cardiovascular Imaging* 2020; 13: 524–526
- [15] Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* 2018; 2: 35
- [16] Tota-Maharaj R, Joshi PH, Budoff MJ et al. Usefulness of regional distribution of coronary artery calcium to improve the prediction of all-cause mortality. *Am J Cardiol* 2015; 115: 1229–1234
- [17] Blaha MJ, Budoff MJ, Tota-Maharaj R et al. Improving the CAC Score by Addition of Regional Measures of Calcium Distribution: Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovascular imaging* 2016; 9: 1407–1416
- [18] Greenland P, Bonow RO, Brundage BH et al. ACCF/AHA 2007 Clinical Expert Consensus Document on Coronary Artery Calcium Scoring By Computed Tomography in Global Cardiovascular Risk Assessment and in Evaluation of Patients With Chest Pain: A Report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) Developed in Collaboration With the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. *Journal of the American College of Cardiology* 2007; 49: 378–402
- [19] Datta D. blandr: a Bland-Altman Method Comparison package for R. Zenodo doi 2017: 10
- [20] Shahzad R, van Walsum T, Schaap M et al. Vessel Specific Coronary Artery Calcium Scoring: An Automatic System. *Academic radiology* 2013; 20: 1–9
- [21] Blaha MJ, Mortensen MB, Kianoush S et al. Coronary Artery Calcium Scoring. *JACC: Cardiovascular Imaging* 2017; 10: 923–937
- [22] Willemink MJ, Vliegenthart R, Takx RA et al. Coronary artery calcification scoring with state-of-the-art CT scanners from different vendors has substantial effect on risk classification. *Radiology* 2014; 273: 695–702
- [23] Weininger M, Ritz KS, Schoepf UJ et al. Interplatform reproducibility of CT coronary calcium scoring software. *Radiology* 2012; 265: 70–77
- [24] van Assen M, Martin SS, Varga-Szemes A et al. Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: A validation study. *European journal of radiology* 2021; 134: 109428
- [25] Lessmann N, Išgum I, Setio AA et al., editors. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. *Medical Imaging 2016: Computer-Aided Diagnosis*; 2016: International Society for Optics and Photonics.
- [26] Velzen SGMv, Lessmann N, Velthuis BK et al. Deep Learning for Automatic Calcium Scoring in CT: Validation Using Multiple Cardiac CT and Chest CT Protocols. *Radiology* 2020; 295: 66–79