Improved bias and reproducibility of coronary artery calcification features using deconvolution

Yingnan Song, Ammar Hoori, Hao Wu, Mani Vembar, Sadeer Al-Kindi, Leslie Ciancibello, James G. Terry, David R. Jacobs Jr, John Jeffrey Carr, and David L. Wilson

Abstract

Purpose: Our long-range goal is to improve whole-heart CT calcium scores by extracting quantitative features from individual calcifications. Here, we perform deconvolution to improve bias/reproducibility of small calcification assessments, which can be degraded at the normal CT calcium score image resolution.

Approach: We analyzed features of individual calcifications on repeated standard (2.5 mm) and thin (1.25 mm) slice scans from QRM-Cardio phantom, cadaver hearts, and CARDIA study participants. Preprocessing to improve the resolution involved of Lucy–Richardson deconvolution with a measured point spread function (PSF) or three-dimensional blind deconvolution in which the PSF was iteratively optimized on high detail structures such as calcifications in images.

Results: Using QRM with inserts having known mg-calcium, we determined that both blind and conventional deconvolution improved mass measurements nearly equally well on standard images. Further, deconvolved thin images gave an excellent recovery of actual mass scores, suggesting that such processing could be our gold standard. For CARDIA images, blind deconvolution greatly improved results on standard slices. Bias across 33 calcifications (without, with deconvolution) was (23%, 9%), (18%, 1%), and (−19%, −1%) for Agatston, volume, and mass scores, respectively. Reproducibility was (0.13, 0.10), (0.12, 0.08), and (0.11, 0.06), respectively. Mass scores were more reproducible than Agatston scores or volume scores. For many other calcification features, blind deconvolution improved reproducibility in 21 out of 24 features. Cadaver images showed similar improvements in bias/reproducibility and slightly better results with a measured PSF.

Conclusions: Deconvolution improves bias and reproducibility of multiple features extracted from individual calcifications in CT calcium score exams. Blind deconvolution is useful for improving feature assessments of coronary calcification in archived datasets.

© 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.10.1.014002]

Keywords: CT calcium score; cardiovascular disease; deconvolution; image processing.

Paper 22080GRRR received Mar. 27, 2022; accepted for publication Dec. 7, 2022; published online Jan. 13, 2023.
1 Introduction

Cardiovascular disease is the most common cause of death in the United States, and coronary artery disease (CAD) is the most common type of heart disease.\textsuperscript{1,2} CT coronary artery calcium (CAC) gives direct evidence of atherosclerotic coronary artery disease, which can be obtained via a fast, reliable, noninvasive, and noncontrast-enhanced examining method. As coronary calcium is easy to reliably detect, it results in an examination with extremely high sensitivity/specificity. The exam is low cost, creating an opportunity for screening individuals in high-risk categories. At our institution (University Hospitals of Cleveland), the CT calcium score exam is currently free, with \(\sim 13,000\) exams conducted annually.

A great number of studies have shown that the CT calcium whole-heart Agatston score predicts risk of adverse cardiovascular events.\textsuperscript{3–9} The Agatston score, as an accurate marker of subclinical coronary artery disease, is more predictive than other single biomarkers, including lipids.\textsuperscript{5,10} There is emerging evidence that shows that regional calcification is important.\textsuperscript{11,12} The whole heart Agatston score is a traditional measure of calcium on a coronary CT calcium scan. However, individual calcifications are important because whole-heart Agatston can easily be dominated by the largest, densest, stable calcifications. From pathobiology and clinical observations, it is likely that small, spotty, low-density calcifications will provide better evidence of disease progression than the whole-heart Agatston score. In fact, a recent study suggested that patients with very high calcification densities \([>1000\text{ Hounsfield Unit (HU)}]\) have reduced risk,\textsuperscript{13} contrary to Agatston, suggesting room for improvement.\textsuperscript{11}

Other common assessments in CT calcium score exams are volume and mass scores.\textsuperscript{14–16} The volume score is the total number of calcified arterial voxels, and the mass score is the accumulation of an actual mineral mass expressed in milligrams,\textsuperscript{17} which has inherent advantages in the presence of partial volume effects. Various studies have showed that the mass score is less variable compared with the Agatston score and volume score.\textsuperscript{18–21} Typical Agatston, volume, and mass scores are given as single numbers for the entire heart. In our work, we are particularly interested in identifying features of individual calcifications (e.g., number of lesions detected in 3D, maximum HU value, maximum mass score, and more), including small calcifications, suggesting a need for corrections aimed at obtaining more accurate and reproducible measurements.\textsuperscript{22} Several papers have identified that feature reproducibility is an important requirement for machine learning.\textsuperscript{23–27}

There is related image processing work. CT images are blurred due to the focal spot size, reconstruction filter, motion of the gantry during sampling of a projection, etc. To address blurring, Liang et al.\textsuperscript{28} applied deconvolution image restoration to reduce the blooming in cardiac CT images, and Rollano-Hijarrubia et al.\textsuperscript{29} claimed that deconvolution restores small high-density structures in micro-CT, which enhanced the visualization of calcification. In a preliminary report, Richards et al. developed a motion point spread function (PSF) based method on stenosis estimation,\textsuperscript{30,31} and Yongpan et al.\textsuperscript{32} proposed using different weight values for the smooth region and edge region during the deblurring process to suppress ringing. There are additional reports of using deconvolution-type processing to improve image quality in CT.\textsuperscript{33–35} More recently, deep-learning-based methods have also been used to improve image sharpness.\textsuperscript{36} Apart from deconvolution, partial volume corrections can be applied on calcifications, wherein the assumption is that voxels at an interface are “averaging” a mixture of a calcification and soft tissue. Recently, we developed a method for partial volume correction,\textsuperscript{37} and Šprem et al.\textsuperscript{38} reported a similar approach.

Rather than targeting visual image quality improvement, our goal is to improve quantification of calcifications. In this study, we analyzed and compared the use of appropriate deconvolution correction methods on CT calcium score exam images obtained from phantoms with known calcification mass, heavily calcified cadaver hearts, and the Coronary Artery Risk Development in Young Adults (CARDIA) cohort study. In experiments, we analyzed the bias and reproducibility of features of individual calcifications with and without deconvolution of standard slice thickness (2.5 mm) scans. As the 3D PSF of a CT imaging system is not always known, we compared conventional deconvolution with known PSF and blind deconvolution, which estimates the PSF from the image itself. In addition to Agatston, mass, and volume of individual
calcifications, we analyzed the reproducibility of various features for potential use in machine learning analysis of the major adverse cardiovascular event risk.

2 Methods

2.1 CT Imaging

To estimate the reproducibility of calcification features, CT imaging of phantoms and cadavers was performed as follows: the QRM-Cardio phantom with known mass values and 10 cadaver hearts were scanned at three different angles (−15 deg, 0 deg, and +15 deg) using Philips IQon spectral CT (Philips Healthcare, Best, The Netherlands) in standard CT mode. The scan protocol used was a standard slice scan (SS, slice thickness: 2.5-mm contiguous, exposure: 55 mAs, and filter type: CB) and a thin slice scan (TS, slice thickness: 0.67-mm contiguous, exposure: 200 mAs, and filter type: CB). The reconstruction method was “cone beam” with a 0.49-mm in-plane voxel size.

We also analyzed 20 participants’ scan/rescan data from CARDIA Y20. The CARDIA dataset recruited young black and white men and women at ages 18 to 30 years in 1985 and followed up at 5-year intervals. At Y20, repeated scans were acquired at about a 5-min interval. Images were acquired at the standard slice scan (SS, slice thickness: 2.5 contiguous, exposure: 100 mAs, filter type: body filter) and the thin slice (TS, slice thickness: 1.25 mm contiguous, exposure: 100 mAs, filter type: body filter) using a similar pixel size (0.68 mm in-plane). Using high quality interpolation (3D interpolation based on cubic convolution), we converted slice thicknesses of all TS scans of the cadaver heart and phantom to be consistent with the CARDIA dataset acquisition (slice thickness: 1.25 mm).

2.2 Calcification Features

As described in the Introduction, we wanted to analyze individual calcifications. We computed traditional features (Agatston, volume, and mass) using methods described in the literature. The Agatston score was calculated according to the maximum HU value of the calcified region (≥ 3 connected voxels over 130 HU) in axial 2D CT images. The total Agatston score was calculated by multiplying a weighting factor with the 2D area of the individual calcified region in each axial slice, with the weighting factor determined by the maximum HU value. The volume score was calculated using the calibration curve generated from the mean HU value of two calibration QRM phantom inserts, which represented the relationship between the HU value of the voxel and calcium concentration. We manually segmented the calcification masks accepting connected voxels over 130 HU and saved the individual calcified region as a binary mask. We determined calibrations for the CT scanner using two calibration calcium inserts in the QRM phantom (shown in Fig. 1). For the mass score assessment, we assumed that the HU value varies linearly with density of hydroxyapatite (HA) with an offset in water inserts and used water equivalent material inserts and 200 mgHA/cm³ calibration inserts to generate the curve. Then, we selected the volume at the center of the calibration inserts and calculated the mean HU value of the voxels to generate calibration factor \( k \). The mass score is evaluated using Eq. (1). To validate our results, we also analyzed calcium mass scores using a commercial software for comparison:

\[
\text{Mass score} = \sum k \times \text{HU}_\text{voxel},
\]

where \( k \) is the calibration factor generated by the two calibration inserts in the phantom and \( \text{HU}_\text{voxel} \) represents the HU value of the voxel selected. In our experiment, calibration factor \( k \) in the phantom images was 0.71 (mgHA/cm³)/HU for 120 kVp images.

We also evaluated Individual Calcification MORphological and mass featurEs (ICmore). ICmore is a comprehensive analysis of the individual calcification in the CT calcium score exam, which includes HU value-related features such as the mean/median HU value and morphological features such as volume and first moment. We divided these into two groups: accumulated heart
and artery (AHA) and 3D calcification features. AHA features are the analysis of the whole heart, such as the whole-heart Agatston, mass and volume scores, whereas 3D features look into individual calcifications, such as mean, max, and standard deviation of the mass score. Details can be found in Fig. 10.

2.3 PSF Measurement and Deconvolution

The generation of CT images is modeled as a linear space-invariant system as

$$I(x, y, z) = h(x, y, z) * I_{\text{origin}}(x, y, z) + n(x, y, z),$$

(2)

where $I(x, y, z)$ is the output image; $h(x, y, z)$ is the PSF of the system; $I_{\text{origin}}(x, y, z)$ represents the real image structure, which is the input of the system; and $n(x, y, z)$ is additive noise.

This operation in Eq. (2) blurred the image details, especially the edges and small objects, thus reducing the bias of the calcium mass score calculation. To recover the real image, we need to estimate the additive noise and the PSF. We measured the additive noise by selecting an area composed of the same material expected to be homogeneous on the phantom. To measure the 3D PSF, we used tiny plastic sphere beads (diameter: 0.2 mm), which are smaller than one voxel and were embedded in OCT gel as phantom to provide adequate contrast. We acquired the PSF from nine beads and averaged them. Both SS (slice thickness: 2.5 mm) and TS (slice thickness: 1.25 mm) scans were obtained, and to match slice thicknesses, we interpolated the SS volume to TS using 3D interpolation based on cubic convolution. To measure PSFs, we manually selected beads with a bounding box, subtracted the background, and fit a 10-parameter 3D Gaussian model ($3 \sigma$’s; scaling parameter; $x, y$, and $z$ offsets and peak intensities) by minimizing the squared error.

The Lucy–Richardson method is an iterative procedure for recovering an underlying image that has been blurred by a known PSF.30,31 This algorithm uses both the PSF and noise characteristics as priors in the processing, enabling high-contrast images with reduced noise compared with other deconvolution methods. We used the damped Richardson–Lucy deconvolution algorithm implemented in MATLAB (MathWorks). The 3D PSF consisted of a $6.24 \times 6.24 \times 8.71 \text{ mm}^3$ noise-free 3D array generated from the Gaussian PSF model. In processing, we set some adjustable parameters (values) as follows: the maximum number of

![Fig. 1](QRM_phantom_datasheet.png)

**Fig. 1** QRM phantom datasheet. (a) QRM phantom with nine calcification inserts of three sizes (1, 3, and 5 mm in diameter) and three densities (200, 400, 800 mg HA/cm³) and with two calibration inserts (water and 200 mg HA/cm³). More information can be found on the company’s website.40 Densities and sizes of inserts were given in the datasheet, so we can determine gold standard measurements. The density, volume, and mass scores were given, so we can infer the theoretical HU value according to the calibration curve. Theoretical HU values were 302, 623, and 1268 HU, respectively. (b) Surface plot of a CT image slice (standard slice thickness) containing the calcification inserts. Even though the calcification density was fixed along a radius, the smaller inserts showed significantly less peak signal.
iterations (80 times), the threshold for damping (0 HU), and subsampling (exact size). In addition, we provided the algorithm with an “additive noise array” obtained from calcification-free regions. Similar to Yongpan et al.32 who used weight values for deblurring, we created a weighted array consisting of ones in a relatively large region around each calcification and zeros elsewhere. Anisotropic diffusion filtering was applied after deconvolution to depress the noise. By running the diffusion filter with a 3D edge-seeking diffusion coefficient for a certain number of iterations (five times), the image can be evolved toward a piecewise constant image with the boundaries between the constant components being detected as edges. We applied 3D Lucy–Richardson deconvolution corrections using a measured PSF on the phantom and cadaver heart volume of SS images (standard slice thickness after deconvolution, SSD) and TS images (thin slice thickness after deconvolution, TSD). We used a 3D blind damped Lucy–Richardson deconvolution algorithm on CARDIA participants’ volume on SS images (standard slice thickness after blind deconvolution, SSBD). It utilizes the maximum likelihood algorithm and the initial estimate of the PSF. The initial estimate for the PSF was set to the same value as that which we measured above.

2.4 Evaluation Method

We analyzed the QRM phantom with nine calcification inserts of three sizes (1, 3, 5 mm in diameter) and three densities (200, 400, 800 mgHA/cm³) and used two calibration inserts (water and 200 mgHA/cm³). Densities and sizes of inserts were given in the datasheet. We also analyzed 92 individual calcifications in 10 cadaver hearts and 33 individual calcifications in 20 CARDIA participants. For the CT calcium score, the bias was defined as the average percent signed difference {mean [measurement − gold standard]/gold standard}, and reproducibility was defined as the coefficient of variation in the angled or repeated scans. We easily obtained the gold standard in the QRM phantom with given values and created ground truth data in the cadaver hearts study and CARDIA participants; details are given below in Results. In addition, we analyzed the reproducibility of ICmore across repeated scans in CARDIA participants using the intraclass correlation coefficient (ICC).

3 Results

A significant reduction in peak intensity was observed when imaging calcification inserts in the QRM phantom (Fig. 1). The effect was profound on the smallest calcification insert (1 mm diameter, 1 mm height). For the smallest 800 mgHA/cm³ insert, the peak HU value (152 HU) was reduced by 90% compared with the actual HU value based on hydroxyapatite concentration (1268 HU). The HU value barely exceeded the standard threshold for calcification detection (i.e., 130 HU).

For PSF measurements, we found insignificant changes in parameters in different locations and averaged the parameters across nine beads to get an average PSF. From 3D PSF measurements obtained in SS and TS scans, we got $\sigma_x = \sigma_y = 0.58 \pm 0.08$ mm in the transverse plane, with $\sigma_z$ of $1.5 \pm 0.5$ mm (slice thickness: 2.5 mm) and $0.84 \pm 0.34$ mm (slice thickness: 1.25 mm), respectively. There was very good agreement between bead measurements obtained with a clinical scan in the transverse plane to those reported previously using the Catphan 600 phantom.41

We obtained images of the QRM phantom and visually evaluated the effect of deconvolution (Fig. 2). We took the TS as the reference in this set of images, and we could see significant brightness reduction in SS especially in the smallest calcium inserts with high density (red circle). However, after correction, SSBD and SSD recovered image contrast significantly. In the corrected clinical scans, the center HU value increased by 68% (164 to 276 HU) and 74% (164 to 286 HU) after SSBD and SSD, respectively, closer to the actual value (1268 HU). The peak HU difference (286 and 291 HU) between TS and SSD was only 2% in this insert.

Then we quantitatively compared calcification mass scores of the QRM phantom obtained from SS and TS scans, with and without deconvolution (Fig. 3, Table S1 in the
Supplementary Material). The actual “ground truth” phantom mass scores (dashed lines) were obtained from the concentration of hydroxyapatite (HA) and the volume of the inserts. Figure 3(a) describes the clinically relevant density inserts in different sizes, and Fig. 3(b) describes the subclinical inserts with different densities. SSD correction improved bias from 13% to 8% and reproducibility improved from 0.12 to 0.04 across the angles for all cases (Sec. 2.4). All small inserts in the QRM phantom were detected by our standard criterion and provided reasonable results compared with the ground truth. TSD yielded excellent bias results across all measurements with values within about 2.1% of the ground truth. Reproducibility showed similar results as the coefficient of variation reduced from 0.18 to 0.14 after correction. Obtaining the true mass score of each individual calcification within the hearts is a laborious task. As a result, in the cadaver heart experiments, we assumed TSD to be the “gold standard” mass score, even if it was not the real physical value, as it was the most accurate practical

Fig. 2 Improvements in image quality with deconvolution (phantom). We compared the 0 deg images at (a) TS thickness (TS), (b) SS thickness (SS), (c) SSBD, (d) SSD, and (e) TSD. In each image, there are left and right sets of three insets containing 400 and 800 mgHA/cm³, respectively. In the insets, we zoom the smallest insert (1 mm in diameter and height) to illustrate the effect of deconvolution. The maximum HU value in the inset is shown. After deconvolution of the standard slice images (SSBD and SSD), images are close to the TS image.

Fig. 3 Comparisons with actual QRM-Cardio phantom values. Evaluations TSD, SS thickness, SSD, and commercial software were compared with actual mass scores (dashed line). Error bars were from the scan with angles. Ag stands for the Agatston score. (a) Results in clinically relevant calcification inserts and (b) results in smallest calcification inserts. In (a), TSD excellently agreed with the actual, leading us to use this as the gold standard in heart imaging. Deconvolution significantly improved results from standard images, especially in (b), and we can see that small calcification inserts can be detected and giving reasonable results after deconvolution.
measurement. Similarly, for the Agatston score (Fig. S2 in the Supplementary Material), deconvolution processing improved the bias from 11% to 7% and reproducibility improved from 0.44 to 0.32.

We scanned and analyzed calcifications in 10 cadaver hearts, and most of them were isolated spotty calcifications with a mass score of less than 20 mg calcium. As for the cadaver images (Fig. 4), we could visually see that SSD corrected the calcifications to be more similar to TS scan, and contrast was restored as the edge of zoomed calcification became much clearer. In Fig. 5 and Table S2 in the Supplementary Material, we selected four calcifications from a cadaver heart (CA1-4) to make detailed comparisons of measurements with and without correction. In SS, scores were underestimated compared with the gold standard, but after correction, SSD results were the best. The variation in angled measurements was reduced with SSD in some instances. In Fig. 6, we presented a modified Bland–Altman plot of all 92 calcification mass score evaluations. In all cases, we compared measurements with the gold standard measurement (TSD). The plot shows bias in a measurement when the average horizontal curve is different from the axis at zero; shows the spread of measurements, indicating precision; and shows reproducibility as each datum includes the mean and standard deviation across measurements at different angles. We also include SSBD results in this plot. Considering the bias and precision results, methods for processing can be ordered as SSD > SSBD > commercial software ≈ SS using our analysis software. For example, mean differences were −2.3, −4.32, −9.42, and −10.59 mg calcium, respectively. Similarly, mean standard deviations were 3.96, 4.53, 8.35 and 9.07 mg calcium, respectively. Similar results were obtained for the Agatston score (Fig. S3 in the

Fig. 4 Improvements in image quality with deconvolution (cadaver heart). We showed 0 deg images at (a) TS thickness, (b) SS thickness, (c) SSBD, (d) SSD, and (e) TSD correction. The subimages with the red border were the zoomed-in calcifications in the circled locations. The mass scores were 15.43, 11.59, 13.23, 14.52, and 18.85, in units of mg calcium, respectively. After deconvolution, SSBD and SSD were much improved compared with SS.

Fig. 5 Comparisons with thin slice deconvolution cadaver values. Image conditions (SS thickness, SSD and commercial software) are compared with TSD (dashed). Deconvolution significantly improved results from standard images.
Fig. 6 Modified Bland–Altman plot of the calcium mass score in 10 cadaver hearts. The X-axis is the reference mass score, which is TSD results. The Y-axis is the difference between the measurement and reference. The solid black line describes the mean difference over the measurement in the figure, and the dashed line describes the limits of agreement. Ideal measurements come with a mean difference close to 0 and narrow range of limits of agreement. SSD performs better than the commercial software, improving the bias by 78% and reproducibility by 52% compared with SS thickness, and SSBD shows similar performance compared with the SSD. Large calcifications are less reproducible and underestimated a lot even in the corrected analysis.

Fig. 7 Patient level bias/reproducibility of CARDIA mass scores following blind deconvolution. Image conditions (SS thickness, SSBD, TS thickness) are compared with TSBD (dashed). Error bars are from scan with scan-rescan. According to the degree of coronary artery calcification, we divide our 20 participants into three groups [(a) Agatston 1–100, (b) 101–400, and (c) above 400]. Similar to cadaver heart results, SSBD improved results from SS as the average percent signed difference reduced from −19% to −1%. Also, the averaged coefficient of variation reduced from 0.11 to 0.06.
Mean differences were $-2.61$, $-2.82$, $8.94$, and $9.03$, respectively, and mean standard deviations were $11.41$, $13.75$, $14.58$, and $16.41$, respectively.

In CARDIA images, we found that SSBD improved bias (Fig. 7, Table S3 in the Supplementary Material, Figs. 8–10, Fig. S4 in the Supplementary Material) compared with our reference TSBD (Thin Slice after blind deconvolution). For Agatston, volume, and mass scores, the calculated average percent signed difference were (23%, 9%), (18%, 1%), and ($-19\%$, $-1\%$), respectively for SS and SSBD. For Agatston, volume, and mass scores, there were insignificant differences between measurements in SSBD and the reference, TSBD, (one-sample t-test, $p = 0.47$, $p = 0.56$, and $p = 0.63$, respectively), suggesting good accuracy of results after deconvolution. Analysis of Agatston, volume, and mass scores showed the superiority of blind deconvolution compared with standard images (SSBD versus SS) ($p \sim 0.01$, one-sided paired t-test). For reproducibility, we calculated the average coefficient of variation, results were (0.13, 0.1), (0.12, 0.08), and (0.11, 0.06), for SS and SSBD, for Agatston, volume, and mass scores, respectively, showing the most reproducible assessment was mass score following deconvolution. Similar results were in the cadaver heart: the mass score showed an improved bias from 26% to 3% and reproducibility from 0.14 to 0.08. We also investigated the reproducibility of calcification features in CARDIA scan-rescan. We observed improvement in calcification number reproducibility using blind deconvolution (Fig. 9). Deconvolution reduced differences in scan–rescan from 9 to 2. We analyzed 24 features from our ICmore, and a bar plot of ICC showed that the correlation of ICmore improved after blind deconvolution in CARDIA scan–rescan (Fig. 10).

Fig. 8 Modified Bland–Altman plot of the calcium score in CARDIA participants. We analyzed 33 calcifications in TS thickness (1.25 mm) and SS thickness (2.5 mm) reconstructions and compared SS and SSBD with TSBD. Similar to the cadaver heart, SSBD reduces limits of agreement. The mean difference reduces from $-1.82$ to $-0.27$ mg calcium, with the averaged standard deviation being comparable, from 4.2 to 6.5 mg calcium.

Fig. 9 Deconvolution improves the reproducibility of the calcification number detected in CARDIA participants. This figure listed the number of calcifications that we detected in the CARDIA dataset for these 20 participants. Each participant had repeated scans; the number of calcifications detected were plotted as paired bars. (a) SS thickness analysis and (b) SSBD. Each arrow indicated disagreement by one calcification. Deconvolution reduced differences from 9 to 2.

### Supplementary Material

Mean differences were $-2.61$, $-2.82$, 8.94, and 9.03, respectively, and mean standard deviations were 11.41, 13.75, 14.58, and 16.41, respectively.

In CARDIA images, we found that SSBD improved bias (Fig. 7, Table S3 in the Supplementary Material, Figs. 8–10, Fig. S4 in the Supplementary Material) compared with our reference TSBD (Thin Slice after blind deconvolution). For Agatston, volume, and mass scores, the calculated average percent signed difference were (23%, 9%), (18%, 1%), and ($-19\%$, $-1\%$), respectively for SS and SSBD. For Agatston, volume, and mass scores, there were insignificant differences between measurements in SSBD and the reference, TSBD, (one-sample t-test, $p = 0.47$, $p = 0.56$, and $p = 0.63$, respectively), suggesting good accuracy of results after deconvolution. Analysis of Agatston, volume, and mass scores showed the superiority of blind deconvolution compared with standard images (SSBD versus SS) ($p \sim 0.01$, one-sided paired t-test). For reproducibility, we calculated the average coefficient of variation, results were (0.13, 0.1), (0.12, 0.08), and (0.11, 0.06), for SS and SSBD, for Agatston, volume, and mass scores, respectively, showing the most reproducible assessment was mass score following deconvolution. Similar results were in the cadaver heart: the mass score showed an improved bias from 26% to 3% and reproducibility from 0.14 to 0.08. We also investigated the reproducibility of calcification features in CARDIA scan-rescan. We observed improvement in calcification number reproducibility using blind deconvolution (Fig. 9). Deconvolution reduced differences in scan–rescan from 9 to 2. We analyzed 24 features from our ICmore, and a bar plot of ICC showed that the correlation of ICmore improved after blind deconvolution in CARDIA scan–rescan (Fig. 10).
Discussion

In our study, we reported that deconvolution improves the bias and reproducibility of multiple features extracted from individual calcifications in the CT calcium score exam, especially for small calcifications, which have been reported to have higher risk of acute disease. Further, blind deconvolution performed well compared with conventional Lucy-Richardson deconvolution. Therefore, utilizing this technique may be beneficial for predicting adverse cardiovascular events as it provides a robust and accurate evaluation of clinically relevant and preclinical calcifications.

Feature reproducibility is a requirement for rigorous machine learning results. To address this issue, several studies have investigated the stability of feature selection algorithms, measuring the robustness of the selected features in the data. In the case of our results, we investigated features of the individual coronary artery calcification and traditional CT calcium score. The mass score performed the best in traditional measurements; individual calcification features such as number of lesions detected significantly improved reproducibility after blind deconvolution. Blind deconvolution improved bias and reproducibility in our study across the features. We conclude that, for archived images in which PSFs are unavailable, 3D blind deconvolution is a useful preprocessing step for improved radiomics assessments of CT calcium score images.

We elegantly created ground truth data in the CARDIA dataset. We selected the most approximate evaluation to the ground truth in the QRM-Cardio phantom, the thin slice thickness after 3D deconvolution correction. We modified the weighting factor based on the calcification volume to calculate “AgatstonScore3D” and “Area2D,” in which the latter comes from the axial view areas across the calcification. The item “stat” refers to the max, min, median, mean, and standard deviation of a feature. The item “per calcium” is a set of individual calcification analyses for a participant. The vertical dashed line shows the threshold of 0.7; any feature across this line shows adequate reproducibility. Blind deconvolution improved ICC in 21 out of 24 features. With blind deconvolution, 22 features were above the threshold.

Fig. 10 Reproducibility of IC more features in the CARDIA repeated scan with and without blind deconvolution. For 20 participants, we analyzed 24 features that were divided into AHA and 3D calcification. The bar plot of ICC showed the correlation of the calcification feature in CARDIA scan-rescan for SS thickness and SSBD. We modified the weighting factor based on the calcification volume to calculate “AgatstonScore3D” and “Area2D,” in which the latter comes from the axial view areas across the calcification. The item “stat” refers to the max, min, median, mean, and standard deviation of a feature. The item “per calcium” is a set of individual calcification analyses for a participant. The vertical dashed line shows the threshold of 0.7; any feature across this line shows adequate reproducibility. Blind deconvolution improved ICC in 21 out of 24 features. With blind deconvolution, 22 features were above the threshold.

4 Discussion

In our study, we reported that deconvolution improves the bias and reproducibility of multiple features extracted from individual calcifications in the CT calcium score exam, especially for small calcifications, which have been reported to have higher risk of acute disease. Further, blind deconvolution performed well compared with conventional Lucy-Richardson deconvolution. Therefore, utilizing this technique may be beneficial for predicting adverse cardiovascular events as it provides a robust and accurate evaluation of clinically relevant and preclinical calcifications.

Feature reproducibility is a requirement for rigorous machine learning results. To address this issue, several studies have investigated the stability of feature selection algorithms, measuring the robustness of the selected features in the data. In the case of our results, we investigated features of the individual coronary artery calcification and traditional CT calcium score. The mass score performed the best in traditional measurements; individual calcification features such as number of lesions detected significantly improved reproducibility after blind deconvolution. Blind deconvolution improved bias and reproducibility in our study across the features. We conclude that, for archived images in which PSFs are unavailable, 3D blind deconvolution is a useful preprocessing step for improved radiomics assessments of CT calcium score images.

We elegantly created ground truth data in the CARDIA dataset. We selected the most approximate evaluation to the ground truth in the QRM-Cardio phantom, the thin slice thickness after 3D deconvolution correction. Also, we conclude that blind deconvolution showed similar performance to the 3D deconvolution based on the measured PSF, proving the assumption that blind deconvolution on thin slice thickness is a reliable standard when analyzing CARDIA participants.
Disclosure
No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments
This project was partially funded by a research grant from Philips Healthcare (Exhibit B-11Y1-Coronary Calcifications). This research is a collaboration between Case Western Reserve University and University Hospitals of Cleveland. The authors give a special thanks to Mani Vembar for discussion. The Coronary Artery Risk Development in Young Adults Study (CARDIA) is supported by contracts HHSN268201800003I, HHSN268201800004I, HHSN268201800005I, HHSN268201800006I, and HHSN-268201800007I from the National Heart, Lung, and Blood Institute (NHLBI). The analysis of CT data was in part supported by R01-HL098445, R44-HL156811, and R01-HL143484. Hao Wu was supported by the Interdisciplinary Biomedical Imaging Training Program, NIH T32EB007509, administered by the Department of Biomedical Engineering, Case Western Reserve University. This report is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References
1. S. S. Virani et al., “Heart disease and stroke statistics—2020 update: a report from the American Heart Association,” Circulation 141(9), e139–e596 (2020).
2. R. Hajar, “Risk factors for coronary artery disease: historical perspectives,” Heart Views Off. J. Gulf Heart Assoc. 18(3), 109–114 (2017).
3. M. J. Budoff and K. M. Gul, “Expert review on coronary calcium,” Vasc. Health Risk Manag. 4(2), 315–324 (2008).
4. P. Greenland et al., “ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology foundation clinical expert consensus task force (ACCF/AHA writing committee to update the 2000 expert consensus document on electron beam computed tomography) developed in collaboration with the society of atherosclerosis imaging and prevention and the society of cardiovascular computed tomography,” J. Am. Coll. Cardiol. 49(3), 378–402 (2007).
5. H. S. Hecht, “Coronary artery calcium scanning: the key to the primary prevention of coronary artery disease,” Endocrinol. Metab. Clin. North Am. 43(4), 893–911 (2014).
6. M. J. Blaha et al., “Providing evidence for subclinical CVD in risk assessment,” Glob. Heart 11(3), 275–285 (2016).
7. G. Pugliese et al., “The dark and bright side of atherosclerotic calcification,” Atherosclerosis 238(2), 220–230 (2015).
8. J. J. Carr et al., “Association of coronary artery calcium in adults aged 32 to 46 years with incident coronary heart disease and death,” JAMA Cardiol. 2(4), 391 (2017).
9. S. Agarwal et al., “Coronary calcium score predicts cardiovascular mortality in diabetes: diabetes heart study,” Diabetes Care 36(4), 972–977 (2013).
10. S. S. Martin et al., “Dyslipidemia, coronary artery calcium, and incident atherosclerotic cardiovascular disease,” Circulation 129(1), 77–86 (2014).
11. C. Thilo et al., “Correlation of regional distribution and morphological pattern of calcification at CT coronary artery calcium scoring with non-calcified plaque formation and stenosis,” Eur. Radiol. 20(4), 855–861 (2010).
12. E. R. Brown et al., “Coronary calcium coverage score: determination, correlates, and predictive accuracy in the multi-ethnic study of atherosclerosis1,” Radiology 247(3), 669–675 (2008).
13. A. R. van Rosendaal et al., “Association of high-density calcified 1K plaque with risk of acute coronary syndrome,” JAMA Cardiol. 5(3), 282–290 (2020).
14. T. Q. Callister et al., “Coronary artery disease: improved reproducibility of calcium scoring with an electron-beam CT volumetric method,” Radiology 208(3), 807–814 (1998).
15. R. Detrano et al., “Accuracy of quantifying coronary hydroxyapatite with electron beam tomography,” *Invest. Radiol.* **29**(8), 733–738 (1994).
16. S. Ulzheimer and W. A. Kalender, “Assessment of calcium scoring performance in cardiac computed tomography,” *Eur. Radiol.* **13**(3), 484–497 (2003).
17. C. Hong, K. T. Bae, and T. K. Pilgram, “Coronary artery calcium: accuracy and reproducibility of measurements with multi-detector row CT—assessment of effects of different thresholds and quantification methods,” *Radiology* **227**(3), 795–801 (2003).
18. C. H. McCollough et al., “Coronary artery calcium: a multi-institutional, multimufacturer international standard for quantification at cardiac CT,” *Radiology* **243**(2), 527–538 (2007).
19. U. Hoffmann et al., “Evidence for lower variability of coronary artery calcium mineral mass measurements by multi-detector computed tomography in a community-based cohort—Consequences for progression studies,” *Eur. J. Radiol.* **57**(3), 396–402 (2006).
20. J. Horiguchi et al., “Electron beam CT versus 16-MDCT on the variability of repeated coronary artery calcium measurements in a variable heart rate phantom,” *AJR Am. J. Roentgenol.* **185**(4), 995–1000 (2005).
21. C. Hong et al., “Coronary artery calcium quantification at multi-detector row CT: influence of heart rate and measurement methods on interacquisition variability initial experience,” *Radiology* **228**(1), 95–100 (2003).
22. Y. Song et al., “Improved reproducibility of CT calcium score using blind deconvolution,” *Proc. SPIE* **11600**, 116000V (2021).
23. J. E. Park et al., “Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives,” *Korean J. Radiol.* **20**(7), 1124–1137 (2019).
24. B. Zhao et al., “Reproducibility of radiomics for deciphering tumor phenotype with imaging,” *Sci. Rep.* **6**(1), 23428 (2016).
25. P. Leo et al., “Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images,” *J. Med. Imaging* **3**(4), 047502 (2016).
26. M. Schwier et al., “Repeatability of multiparametric prostate MRI radiomics features,” *Sci. Rep.* **10**(1), 9441 (2019).
27. H. Merisaari et al., “Repeatability of radiomics and machine learning for DWI: short-term repeatability study of 112 patients with prostate cancer,” *Magn. Reson. Med.* **83**(6), 2293–2309 (2020).
28. Z. Liang et al., “Calcium de-blooming in coronary CT image,” in *IEEE 7th Int. Symp. BioInf. and BioEng.*., pp. 257–262 (2007).
29. E. Rollano-Hijarrubia, R. Manniesing, and W. J. Niessen, “Selective deblurring for improved calciumification visualization and quantification in carotid CT angiography: validation using Micro-CT,” *IEEE Trans. Med. Imaging* **28**(3), 446–453 (2009).
30. W. H. Richardson, “Bayesian-based iterative method of image restoration*,” *JOSA* **62**(1), 55–59 (1972).
31. L. B. Lucy, “An iterative technique for the rectification of observed distributions,” *Astron. J.* **79**, 745 (1974).
32. W. Yongpan et al., “An improved Richardson–Lucy algorithm based on local prior,” *Opt. Laser Technol.* **42**(5), 845–849 (2010).
33. R. Carmi, O. Shapiro, and D. Braunstein, “Resolution enhancement of X-ray CT by spatial and temporal MLEM deconvolution correction,” in *IEEE Symp. Conf. Record Nucl. Sci.*, Vol. 5, pp. 2765–2768 (2004).
34. N. V. Slavine et al., “An iterative deconvolution algorithm for image recovery in clinical CT: a phantom study,” *Phys. Med.* **31**(8), 903–911 (2015).
35. L. Hehn et al., “Blind deconvolution in model-based iterative reconstruction for CT using a normalized sparsity measure,” *Phys. Med. Aaemathsemcolon Biol.* **64**(21), 215010 (2019).
36. P. Sudhakar et al., “Self-supervised learning for CT deconvolution,” *Proc. SPIE* **11595**, 115953Z (2021).
37. Y. Song et al., “Improved reproducibility of calcium mass score using deconvolution and partial volume correction,” *Proc. SPIE* **10953**, 109531O (2019).
38. J. Šprem et al., “Coronary calcium scoring with partial volume correction in anthropomorphic thorax phantom and screening chest CT images,” *PLoS One* 13(12), e0209318 (2018).
39. M. J. Blaha et al., “Coronary artery calcium scoring: is it time for a change in methodology?” *JACC Cardiovasc. Imaging* 10(8), 923–937 (2017).
40. https://www.qrm.de/en/products/cardiac-calcification-phantom/
41. O. Ozguner et al., “Objective image characterization of a spectral CT scanner with dual-layer detector,” *Phys. Med. Biol.* 63(2), 025027 (2018).
42. P. Raggi et al., “Identification of patients at increased risk of first unheralded acute myocardial infarction by electron-beam computed tomography,” *Circulation* 101(8), 850–855 (2000).
43. J. L. Lustgarten, V. Gopalakrishnan, and S. Visweswaran, “Measuring stability of feature selection in biomedical datasets,” in *AMIA. Annu. Symp. Proc.*, pp. 406–410 (2009).
44. A. Kalousis, J. Prados, and M. Hilario, “Stability of feature selection algorithms: a study on high-dimensional spaces,” *Knowl. Inf. Syst.* 12(1), 95–116 (2007).

Biographies of the authors are not available.