



OPEN Evaluation of an artificial intelligence model for opportunistic Agatston scoring on non-gated chest computed tomography

Suzannah E. McKinney^{1,2✉}, Sarah F. Mercaldo^{1,2,3}, John K. Chin¹, Ankita Ghatak¹, Madeleine A. Halle¹, Sandeep S. Hedgire^{2,3}, Nandini M. Meyersohn^{2,3}, Brian Ghoshhajra^{2,3}, Keith J. Dreyer^{1,2,3}, Mannudeep Kalra^{2,3}, Bernardo C. Bizzo^{1,2,3,5} & James M. Hillis^{1,3,4,5}

The Agatston score is a measure of cardiovascular disease traditionally calculated on cardiac gated computed tomography (CT) of the chest. Cardiac gated CT is resource-intensive, can be hard to access, and involves extra radiation exposure. Artificial intelligence (AI) can be used to opportunistically calculate Agatston scores on non-gated CTs performed for other indications. This retrospective standalone performance assessment compared the accuracy of an AI model (Riverain Technologies ClearRead CT CAC) at calculating Agatston scores on non-gated CTs to both consensus radiologist interpretations on the same CTs and Agatston scores from the original radiology reports of paired cardiac gated CTs. It involved 491 non-contrast CT chest cases acquired at five hospitals in the United States between January 2022 and December 2023; approximately two-thirds had a paired cardiac gated CT. It compared the agreement of Agatston categories (0, 1–99, 100–399 and ≥ 400) using the quadratic weighted Kappa coefficient and the correlation of Agatston scores using the Spearman coefficient. The agreement of Agatston categories between the AI model and ground truth radiologists was 0.959 (95% CI: 0.943 to 0.975); this result was broadly consistent across sex, age group, race, ethnicity and CT scanner manufacturer subgroups. The agreement between the AI model and paired cardiac gated CT was 0.906 (95% CI: 0.882 to 0.927). The correlations of Agatston scores for these two comparisons were 0.975 (95% CI: 0.962 to 0.987) and 0.942 (95% CI: 0.920 to 0.957) respectively. The assessed AI model accurately calculated Agatston scores on non-gated CTs and produced similar scores to paired cardiac gated CTs. Its use could broaden screening for atherosclerotic cardiovascular disease, enabling opportunistic screening on CTs captured for other indications.

The prevalence of cardiovascular disease (CVD) in adults in the United States is 48.6% (127.9 million people)¹. Heart disease and stroke claim more lives each year than cancer and chronic lower respiratory disease combined. Moreover, CVD has huge economic impacts with the annual direct and indirect costs estimated to be \$407.3 billion in the United States in 2018 to 2019. Morbidity and mortality can be significantly reduced through primary and secondary prevention strategies, guided by individualized risk assessments. This approach is particularly important when patients remain asymptomatic², although a challenge for many asymptomatic patients is that they may not have undergone as extensive diagnostic work-up to establish their risk assessment.

Computed tomography (CT) coronary artery calcium (CAC) scoring provides an effective, noninvasive method for predicting CVD. It involves the measurement of Agatston score, which is calculated by multiplying the area of calcifications by a factor derived from their highest attenuation value and then obtaining the sum of these products³. An Agatston score can then be categorized into four groups for disease stratification. This technique has demonstrated its ability to reliably assess risk across various populations and provides particular value for prognostication of patients classified as intermediate risk by traditional risk models^{4,5}. One research group followed a total of 42,224 patients of diverse races and ethnicities for a median of 11.7 years; they found that CAC severity based on Agatston category (0, 1–99, 100–399 and ≥ 400) correlated with risk of all-cause and CVD mortality in all studied race/ethnicity groups⁶. Further, CAC severity was correlated with all cause

¹Mass General Brigham AI, 399 Revolution Dr, Boston, MA 02145, USA. ²Department of Radiology, Massachusetts General Hospital, Boston, MA, USA. ³Harvard Medical School, Boston, MA, USA. ⁴Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. ⁵These authors jointly supervised this work: Bernardo C. Bizzo and James M. Hillis. ✉email: semckinney@mgh.harvard.edu

and CVD mortality in groups who may be poorly represented in patient cohorts informing the standard risk prediction models otherwise recommended by professional bodies⁶. Other large cohort studies have further established Agatston scoring as a valuable tool to assess future CVD risk across different age and ethnicity subgroups^{7–9}.

A key challenge with CT CAC measurement is that it has traditionally involved gating the acquisition of the CT with cardiac rhythm to reduce motion artifact. This approach demands specialized training of radiology technologists and requires patients to undergo a CT for this limited purpose with the associated radiation exposure¹⁰. However, there is the opportunity to use routine, non-gated chest CTs that are obtained for non-cardiac indications, which show CAC but whose presence is not routinely reported¹¹. A 2024 meta-analysis of 108 studies concluded that identification of incidental CAC on non-gated thoracic CT is a useful technique given the occurrence of CAC on non-gated scans was both related to the presence of traditional cardiovascular risk factors (except for smoking status and body mass index) and was predictive of the development of subsequent cardiovascular events and all-cause mortality¹².

A further challenge is the time required by radiology technologists or radiologists to create manual segmentations of calcified regions to calculate an Agatston score. Artificial Intelligence (AI) has been used opportunistically in medical imaging elsewhere¹³ and the use of AI to automate manual segmentation has been proposed to enable opportunistic screening on non-gated chest CT. This study assessed the performance of an AI device (Riverain Technologies ClearRead | CT CAC), which was designed to estimate the Agatston score by accurately segmenting CAC on non-contrast, non-gated chest CT.

This study aimed to assess the performance of this AI device by comparing it to the Agatston score and Agatston category obtained from ground truth radiologists on the same non-gated chest CT. For a subset of CTs, it also compared the device output to the Agatston score obtained from a paired gated CT. This combined approach allowed evaluation of the device compared to both a manual approach on the same non-gated CT and the clinical gold standard on the gated CT.

Methods

Study design

This retrospective standalone model performance study was conducted using radiology cases from five hospitals within the Mass General Brigham (MGB) network between January 2022 and December 2023. It was approved by the MGB Institutional Review Board with waiver of informed consent. It was conducted in accordance with relevant guidelines and regulations including the Health Insurance Portability and Accountability Act (HIPAA). This report followed the Standards for Reporting Diagnostic Accuracy (STARD 2015) reporting guideline. The data from this study include protected health information; some data may be available for research purposes from the corresponding author upon reasonable request.

Case selection

The study cohort was selected using two distinct methods to balance representing all disease severities with reflecting the prevalence of these severities. The first method aimed to ensure representation across the four Agatston categories (0, 1–99, 100–399 and ≥ 400) by taking equal numbers of cases in each category based on original radiology reports. The second method aimed to represent prevalence of CAC amongst the general population by taking cases regardless of the Agatston category. For both methods, the cohort considered non-gated CT chest cases performed in any setting (i.e., inpatient or outpatient). There were no limitations on the original CT chest clinical indication. These cases were obtained from patients at least 30 years of age. Repeat cases from the same patient were removed with only the earliest one maintained.

For the first method, a list of paired cases was created, where patients had received both a gated CT chest and non-gated CT chest within the study time period (average duration [\pm standard deviation] between the two studies of 219.0 ± 175.1 days). The reason for this pairing was that the Agatston score is typically only reported as part of a gated chest CT. The paired approach therefore enabled enrichment of the dataset to ensure the presence of cases across the spectrum of CAC severity; albeit the non-gated chest CT cases were ultimately used in the study. These cases were ordered prospectively by date and then selected consecutively for each of the four Agatston categories, in a balanced manner across the hospitals, until there were 83 cases in each category (i.e., 332 cases total).

For the second method, a list of non-gated chest CT cases was created for patients that did not have a paired gated chest CT case (i.e., could not be selected using the first method). These cases were ordered prospectively by date and then selected consecutively, irrespective of Agatston category, until there were 33–34 cases at each of the five hospitals (i.e., 168 cases total).

Image quality review and series selection

All cases were deidentified and then underwent an image quality review by an American Board of Radiology (ABR)-certified radiologist. Cases were excluded if they contained evidence of prior cardiac surgery, presence of intravenous or oral contrast, presence of para-cardiac metal artifact, presence of significant motion artifact that precluded assessment of CAC or presence of significantly altered thoracic anatomy. A total of 16 cases (2.7% of the cases considered) were excluded (Fig. 1).

At the same time as the image quality review, the radiologist selected the series to be given to the model. To facilitate a range of series slice thicknesses, the radiologist selected up to two soft or standard kernel axial series as available: one that was > 0.5 mm and ≤ 1.5 mm, and one that was > 1.5 mm and ≤ 3 mm. When a case had both series available and they both passed the DICOM metadata review as described below, a single series was subsequently selected using randomization that ensured an even balance between these two slice thickness ranges. The ground truth radiologists and the AI model only received this single selected series. The image

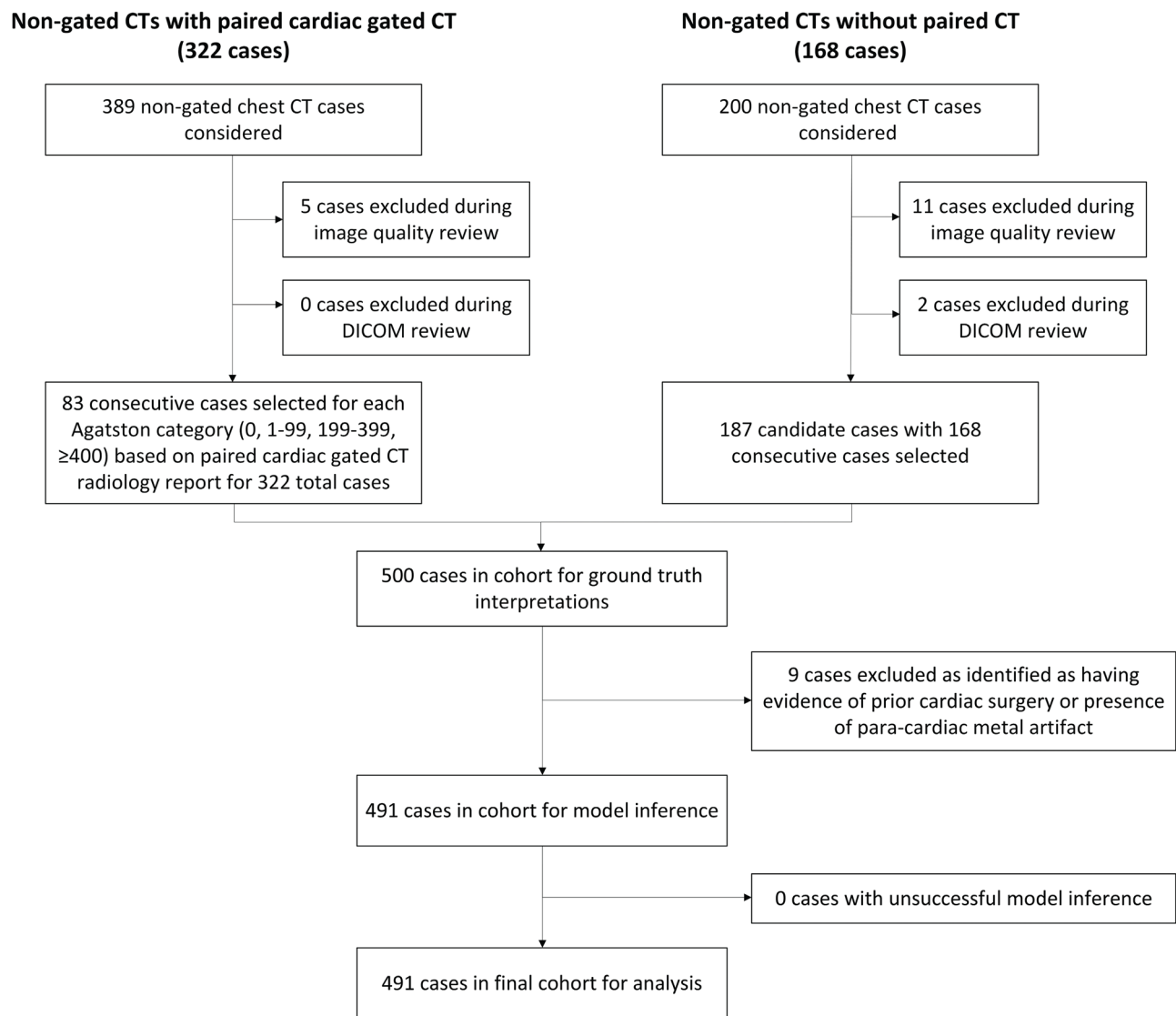


Fig. 1. Cohort diagram.

quality review and series selection process was performed using the FDA-cleared eUnity image visualization software (Version 6 or higher) and an internal web-based annotation system that utilized the REDCap electronic data capture tools hosted at MGB^{14,15}.

DICOM metadata review

A DICOM metadata review was performed after series selection to ensure that the slice thickness was within the appropriate range (>0.5 mm and ≤ 1.5 mm, >1.5 and ≤ 3 mm), the slice thickness was consistent (tolerance of 0.01 mm), the slice interval was consistent (tolerance of 0.01 mm), the slice thickness was equal to or greater than the slice interval (i.e., there were no ‘gaps’ between slices), and the pixel spacing was <1 mm.

Ground truth interpretations

The ground truth interpretations were performed by three ABR-certified radiologists. They independently segmented the regions of CAC using the TeraRecon iNtuition software (version 4.7.0.22–111). As part of this process, they identified the region of CAC as being in the right coronary artery, left anterior descending artery, left circumflex artery and left main coronary artery. The software then calculated the volume of these regions and multiplied them by the Agatston factor to achieve individual Agatston scores for each vessel as well as an overall Agatston score³. The continuous Agatston scores were also categorized using established categories: 0, 1–99, 100–399 and ≥ 400 ¹⁶. We note that Agatston score is traditionally calculated based on the area on 2.5 mm or 3 mm slice thickness series from cardiac gated CT; the volume calculated here involved multiplying axial areas by the slice thickness divided by 3 such that it was scaled to match 3 mm series. In addition, for ease of explanation, we have referred to ‘Agatston score’ for any measurements of CAC that incorporate a volume or

area multiplied by the Agatston factor while recognizing that its strictest definition only applies to the overall score obtained on cardiac gated CT.

The subsequent analyses used the average of the three ground truth radiologists for the continuous Agatston scores (including overall and for each of the four vessels). The analyses for the overall Agatston category used a consensus category, which was the most frequent category amongst the three ground truth radiologists. All cases had a category with at least two ground truth radiologists (i.e., there were no cases where each radiologist recorded a different category). The segmentation masks from each individual radiologist were also used; they were filtered to only include voxels with ≥ 130 Hounsfield units (HU) that would be used for Agatston calculations.

Model inference

The evaluated AI model was version 1.1.1.37 of the Riverain Technologies ClearRead | CT CAC device. In brief, it consisted of a U-Net style encoder–decoder architecture that operates across slices, labeling calcified components into one of several categories, including aortic wall and valve calcifications¹⁷. The model segments CAC and calculates an Agatston score. It was trained on approximately 2,000 chest CTs with varied protocols and acquisition parameters. The model was installed at MGB for use in this study and received only the relevant series from the Digital Imaging and Communications in Medicine (DICOM)-formatted non-gated CT chest cases. It outputted a continuous score for each vessel (right coronary artery, left anterior descending artery, left circumflex artery and left main coronary artery), an overall continuous Agatston score, and an overall Agatston category. The model also provided an intermediate output of the segmentation masks for any identified areas of CAC.

Statistical analysis

The statistical analysis was performed in R (version 4.0.2) on the full analysis set. The outputs used for the statistical analysis included the ground truth radiologist interpretations from the non-gated CT scans, the model outputs from the non-gated CT scans, and the previously reported Agatston scores from paired gated CT scans. In addition, demographic and technical parameters were derived from clinical databases or DICOM fields for subgroup analyses. Any missing data were treated as “Unknown” and no data were imputed. All analyses had 95% confidence intervals (CIs) calculated using bootstrapped intervals with 2,000 resamples.

The predefined primary endpoint was the accuracy of the model in calculating the Agatston category as defined by the quadratic weighted Kappa coefficient when compared with the consensus ground truth radiologist Agatston category. The percentage accuracy (i.e., proportion of cases with matching categories) was also calculated. The sample size had been determined based on powering to ensure the lower bound of the 95% CI of the Kappa coefficient was at least 0.85 using preliminary results.

The predefined analyses for comparing the model to the ground truth radiologist interpretations included the accuracy of the overall and component vessel continuous Agatston scores as defined by the Spearman correlation coefficient. There were also predefined subgroup analyses of the overall Agatston categories and scores for sex, age, race, ethnicity, CT scanner manufacturer and radiation dose protocol; subgroup analyses for slice thickness, days between paired CTs and selection method were added following reviewer feedback. Analyses were performed for comparing the paired gated CT overall Agatston categories and scores to both the model outputs and ground truth radiologist interpretations. These analyses continued to use the quadratic weighted Kappa coefficient and percentage accuracy for Agatston categories, and Spearman correlation coefficient for the Agatston scores. A Bland–Altman analysis was also performed following reviewer feedback.

Additional analyses were conducted to compare the agreement of segmentation masks between the model and ground truth radiologist interpretations using Dice scores. This comparison calculated the Dice score between the device and each of the ground truth radiologists (i.e., device versus radiologist 1, device versus radiologist 2, device versus radiologist 3) as well as between each pair of the ground truth radiologists (i.e., radiologist 1 versus radiologist 2, radiologist 2 versus radiologist 3, radiologist 1 versus radiologist 3). The median Dice score for each set of three measurements was used for a case when all three ground truth radiologists provided a segmentation (i.e., when there was a non-zero Agatston score); the mean Dice score was otherwise used. The overall average Dice score for the device versus ground truth radiologists and between ground truth radiologist pairs was then calculated together with 95% CIs. The difference between the average Dice scores was also calculated. A non-parametric Wilcoxon rank-sum test tested whether the distributions of Dice scores for model versus ground truth radiologists and between ground truth radiologist pairs were significantly different.

Results

The cohort for this study included 491 cases and the model successfully performed inference on all cases (Fig. 1). This cohort included 272 (55.4%) women and 219 (44.6%) men; the mean (standard deviation) age was 65.6 (10.1) years (Table 1). There were 143 cases in Agatston category 0, 125 in Agatston category 1–99, 84 in Agatston category 100–399 and 139 in Agatston category ≥ 400 .

The agreement of the overall Agatston category based on quadratic weighted kappa between the model and consensus ground truth interpretations was 0.959 (95% CI: 0.943 to 0.975) with accuracy of 92.3% (95% CI: 89.9% to 94.6%; Figure 2). The correlation of the overall continuous Agatston score between the model and the average ground truth radiologist interpretations as defined by the Spearman correlation coefficient was 0.975 (95% CI: 0.962 to 0.987). The agreement was broadly consistent across sex, age, ethnicity, race, manufacturer, radiation dose protocol, slice thickness, days between paired CTs and selection method subgroups. All subgroups had a kappa coefficient of at least 0.90 except for the “Other” race subgroup, which had a kappa coefficient of 0.855 for only 11 cases (Table S1). The Bland–Altman analysis demonstrated a mean difference \pm limits of agreement of -12.20 ± 437.84 (Figure S1).

	Agatston score 0	Agatston score 1–99	Agatston score 100–399	Agatston score ≥ 400	Overall
Total cases	143	125	84	139	491
Sex					
Male	42 (29.4%)	50 (40.0%)	39 (46.4%)	88 (63.3%)	219 (44.6%)
Female	101 (70.6%)	75 (60.0%)	45 (53.6%)	51 (36.7%)	272 (55.4%)
Age					
≤ 65 years	95 (66.4%)	70 (56.0%)	40 (47.6%)	46 (33.1%)	251 (51.1%)
> 65 years	48 (33.6%)	55 (44.0%)	44 (52.4%)	93 (66.9%)	240 (48.9%)
Age, Mean ± SD	61.8 ± 9.7	64.2 ± 8.6	66.8 ± 10.0	69.9 ± 10.1	65.6 ± 10.1
Race					
Asian	6 (4.2%)	3 (2.4%)	1 (1.2%)	1 (0.7%)	11 (2.2%)
Black or African American	4 (2.8%)	3 (2.4%)	0 (0.0%)	0 (0.0%)	7 (1.4%)
Other	4 (2.8%)	4 (3.2%)	2 (2.4%)	1 (0.7%)	11 (2.2%)
Unavailable	0 (0.0%)	0 (0.0%)	1 (1.2%)	1 (0.7%)	2 (0.4%)
White	125 (87.4%)	113 (90.4%)	78 (92.9%)	132 (95.0%)	448 (91.2%)
Declined	4 (2.8%)	2 (1.6%)	2 (2.4%)	4 (2.9%)	12 (2.4%)
Ethnicity					
Hispanic	6 (4.2%)	2 (1.6%)	3 (3.6%)	1 (0.7%)	12 (2.4%)
Not Hispanic	129 (90.2%)	116 (92.8%)	76 (90.5%)	131 (94.2%)	452 (92.1%)
Unavailable	5 (3.5%)	6 (4.8%)	3 (3.6%)	6 (4.3%)	20 (4.1%)
Prefer Not to Say/Declined	3 (2.1%)	1 (0.8%)	2 (2.4%)	1 (0.7%)	7 (1.4%)
Manufacturer					
GE Medical Systems	56 (39.2%)	54 (43.2%)	46 (54.8%)	60 (43.2%)	216 (44.0%)
Phillips	5 (3.5%)	2 (1.6%)	5 (6.0%)	8 (5.8%)	20 (4.1%)
Siemens	71 (49.7%)	61 (48.8%)	27 (32.1%)	55 (39.6%)	214 (43.6%)
Toshiba	11 (7.7%)	8 (6.4%)	6 (7.1%)	16 (11.5%)	41 (8.4%)
Radiation dose protocol					
Low dose	74 (51.7%)	69 (55.2%)	54 (64.3%)	72 (51.8%)	269 (54.8%)
Routine dose	69 (48.3%)	56 (44.8%)	30 (35.7%)	67 (48.2%)	222 (45.2%)
Slice thickness					
Thin (> 0.5 mm and ≤ 1.5 mm)	66 (46.2%)	57 (45.6%)	43 (51.2%)	80 (57.6%)	246 (50.1%)
Thick (> 1.5 mm and ≤ 3 mm)	77 (53.8%)	68 (54.4%)	41 (48.8%)	59 (42.4%)	245 (49.9%)
Paired cardiac-gated CT					
No paired gated scan	37 (25.9%)	37 (29.6%)	24 (28.6%)	62 (44.6%)	160 (32.6%)
Paired gated scan	106 (74.1%)	88 (70.4%)	60 (71.4%)	77 (55.4%)	331 (67.4%)
Days between gated and non-gated scans	216.5 ± 184.9	226.4 ± 163.7	206.4 ± 167.8	223.9 ± 181.8	219.0 ± 175.1

Table 1. Demographic and technical breakdown of CT chest cases.

The correlation of Agatston scores for individual arteries showed a Spearman coefficient of 0.951 (95% CI: 0.925 to 0.971) for the right coronary artery, 0.985 (95% CI: 0.978 to 0.991) for the left anterior descending artery, 0.867 (95% CI: 0.824 to 0.904) for the left circumflex artery and 0.791 (95% CI: 0.736 to 0.839) for the left main coronary artery (Fig. 3).

The Dice scores were calculated based on 368 segmentations available from the model and 291, 287 and 288 segmentations from the three ground truth radiologists. Amongst the 302 cases where at least one ground truth radiologist provided a ground truth segmentation, segmentations were available from all three ground truth radiologists for 90.4% of cases, from two ground truth radiologists for 6.0% and from one ground truth radiologist for 3.6%. The average Dice score between the model and ground truth radiologists was 0.892 (95% CI: 0.871 to 0.911) and the average Dice score between ground truth radiologists was 0.910 (95% CI: 0.893 to 0.927). The average difference between Dice scores was -0.008 (95% CI: -0.024 to 0.007), and there was no evidence of a difference between the distributions of Dice scores ($p = 0.12$).

When considering the agreement of the Agatston categories between the model output on the non-gated CT and the reported values for the paired cardiac gated CT, the kappa coefficient was 0.906 (95% CI: 0.882 to 0.927) with accuracy of 79.2% (95% CI: 74.8% to 83.4%; Figure 4). The Spearman correlation coefficient of the Agatston score was 0.942 (95% CI: 0.920 to 0.957). The comparisons between the ground truth radiologist interpretations on the non-gated CT and the reported values for the paired cardiac gated CT had kappa coefficient of 0.907 (95% CI: 0.883 to 0.930), accuracy of 79.5% (95% CI: 75.2% to 83.6%) and Spearman correlation coefficient of 0.941 (95% CI: 0.920 to 0.957; Figure S2).

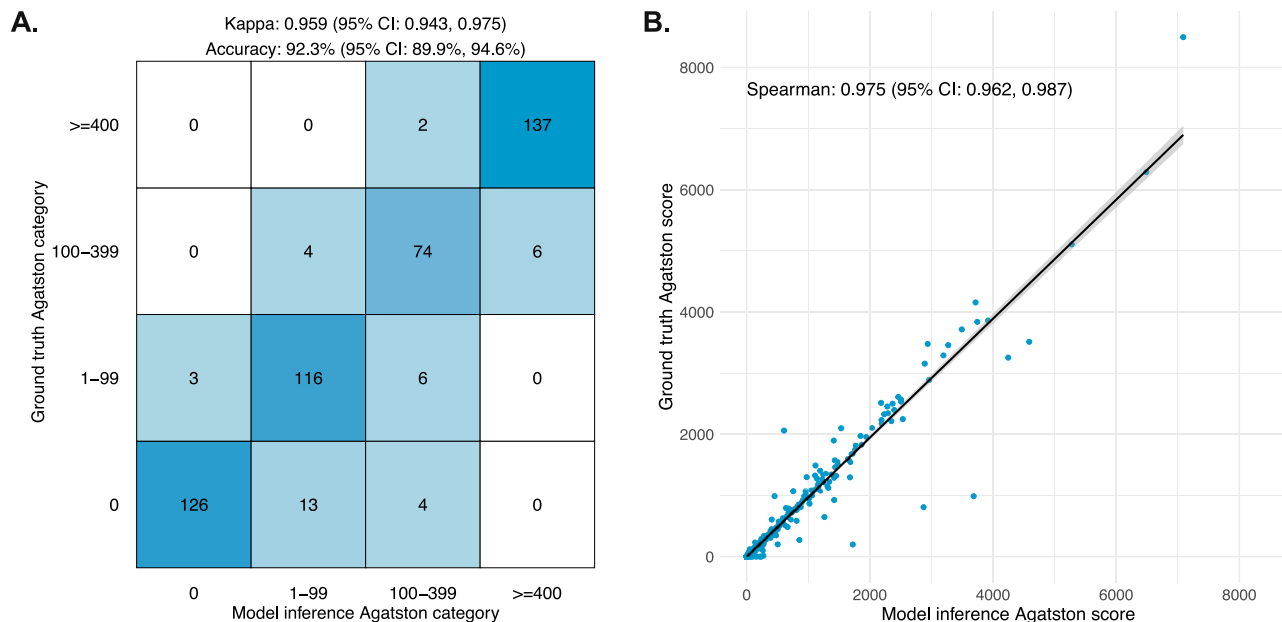


Fig. 2. Comparison of AI model output compared with ground truth radiologists on non-gated CT for the Agatston category (A) and Agatston score (B).

Discussion

This retrospective study assessed the performance of an AI model at quantifying CAC using the Agatston method on non-gated CT. Its accuracy was compared with both ground truth radiologist interpretations of the same non-gated CT and Agatston scores from the original radiology report of paired cardiac gated CTs. The model achieved quadratic weighted kappa coefficients of 0.959 and 0.906 for the respective comparisons for the Agatston categories. It achieved Spearman correlation coefficients of 0.975 and 0.942 for the respective comparisons for the Agatston scores.

These results demonstrate that the model can accurately quantify Agatston scores on non-gated CTs. They show a potential clinical use in screening for incidental coronary artery disease during routine CTs. Importantly, these results also appear to generalize across demographic and technical subgroups, a critical prerequisite for a model to perform consistently across diverse populations⁶. The economic benefits of cardiac gated CTs versus more invasive measures such as cardiac angiography have previously been shown, as have the benefits of risk stratification and modification compared to untreated cardiovascular disease^{10,18}. It may be possible to realize even greater benefits given the higher number of people undergoing routine CTs compared to cardiac gated CTs. This AI model could therefore truly enable opportunistic screening for atherosclerotic cardiovascular disease.

Validation studies that compare AI model quantification of coronary artery calcium on non-gated CT to ground truth results from cardiac gated CT have been reported. The FDA-cleared AVIEW CAC device by Coreline¹⁹ was assessed on a cohort of 567 patients for five Agatston categories (it used the categories 0, 1–10, 11–100, 101–400 and >400); it specifically used low-dose non-gated CT²⁰. This model achieved Cohen's weighted kappa of 0.809 (95% CI: 0.776 to 0.838) with slice thickness 1 mm and 0.776 (95% CI: 0.740 to 0.809) with slice thickness 2.5 mm. A separate model, which involved team members from Bunkerhill who have subsequently received FDA clearance for the iCAC device²¹, achieved kappa coefficients of 0.802, 0.684, 0.644 and 0.583 on a cohort of 303 cases from four different sites²². The current model compared favorably by achieving a kappa coefficient of 0.906 (95% CI: 0.882 to 0.927) for a similar analysis.

Although these previous works generate valuable proof points about the validity of using AI models to calculate Agatston scores from non-gated CT, and therefore the enablement of opportunistic screening on non-gated CT, our study here goes further to demonstrate the viability of opportunistic screening of Agatston scores on non-gated studies in three important ways. Firstly, our study intentionally includes cases from both low-dose and routine-dose protocols, as well as varying slice thicknesses, to assess the model's performance across different types of data it might encounter in real-world scenarios. Secondly, we show stronger correlation between the AI model Agatston score output and consensus ground truth than previously seen. Thirdly, we used as our primary source of ground truth Agatston scores derived from the manual segmentations of three radiologists on a large non-gated dataset of 491 patients.

Although agreement was high all around, there was higher agreement in our study [0.959 (95% CI: 0.943 to 0.975)] between both the AI model output and consensus ground truth than AI model and Agatston categories from paired gated studies [0.906 (95% CI: 0.882 to 0.927)] and between consensus ground truth and Agatston categories from paired gated studies [0.907 (95% CI: 0.883, 0.930)]. Although CAC develops very slowly over time and time periods of <1 year are unlikely to lead to disparate changes between studies, there may be some small impact from the temporal delay between paired studies. Perhaps more importantly, non-gated studies offer a higher spatial resolution than gated due to a thinner typical slice thickness; it is therefore possible that

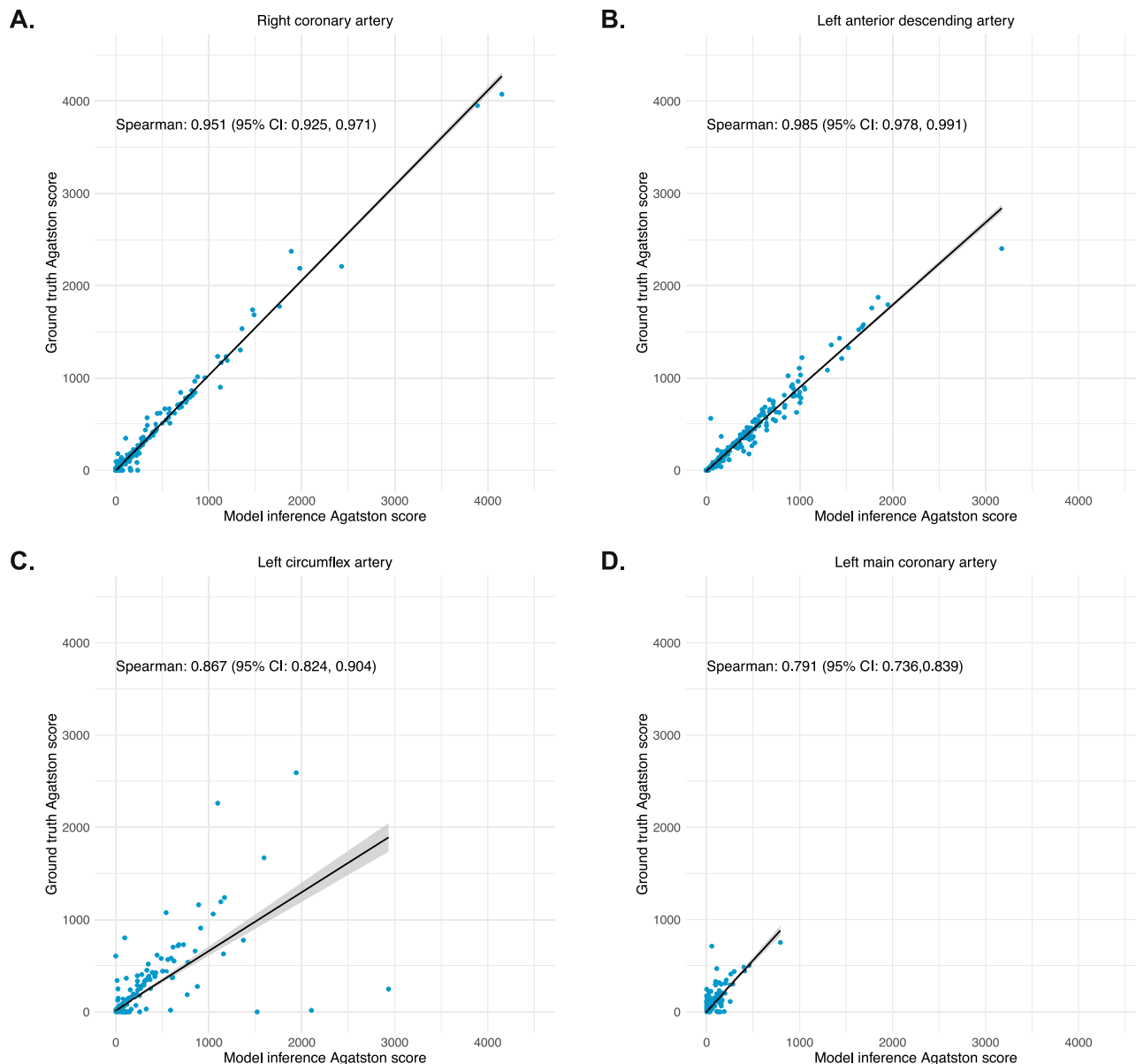


Fig. 3. Scatterplots comparing AI model output with ground truth radiologists on non-gated CT for the Agatston score for each individual vessel including the right coronary artery (**A**), left anterior descending artery (**B**), left circumflex artery (**C**) and left main coronary artery (**D**).

non-gated studies detect calcium that is not seen in gated studies. However, non-gated scans generally contain more noise and may contain cardiac motion that obfuscates calcium measurement, both of which can make calcium quantification more challenging. The current AI model determined the Agatston score based on the HU values for individual voxels within the study (i.e., the Agatston score would be less if the motion artifact caused noise such that the HU values and the corresponding density factor for Agatston score calculation decreased). A potential future consideration is to use an AI model to correct the HU values for cardiac motion.

At a coronary vessel level, the model showed better performance for the right coronary artery and left anterior descending artery (Spearman correlation coefficient of 0.951 and 0.985 respectively) than for the left circumflex artery and left main coronary artery (0.867 and 0.791 respectively). A key possible reason is that the left circumflex artery and left main coronary artery are more susceptible to cardiac motion and therefore more challenging to segment. The left circumflex artery (LCX) and left main coronary artery (LM) are particularly affected by respiratory and cardiac motion due to their lateral and anterior positions, with the LCX being more tortuous and susceptible to motion artifacts. Additionally, the LCX's smaller, more curved vessels experience greater displacement from diaphragmatic motion, while the LM is influenced by the significant movement of the heart's base during the cardiac cycle. In addition, the delineation of the left circumflex artery and the left main coronary artery may not be as clear, which may cause inconsistencies between the ground truth radiologists and the model in attributing calcium to either of these vessels. This situation may reflect why the Agatston

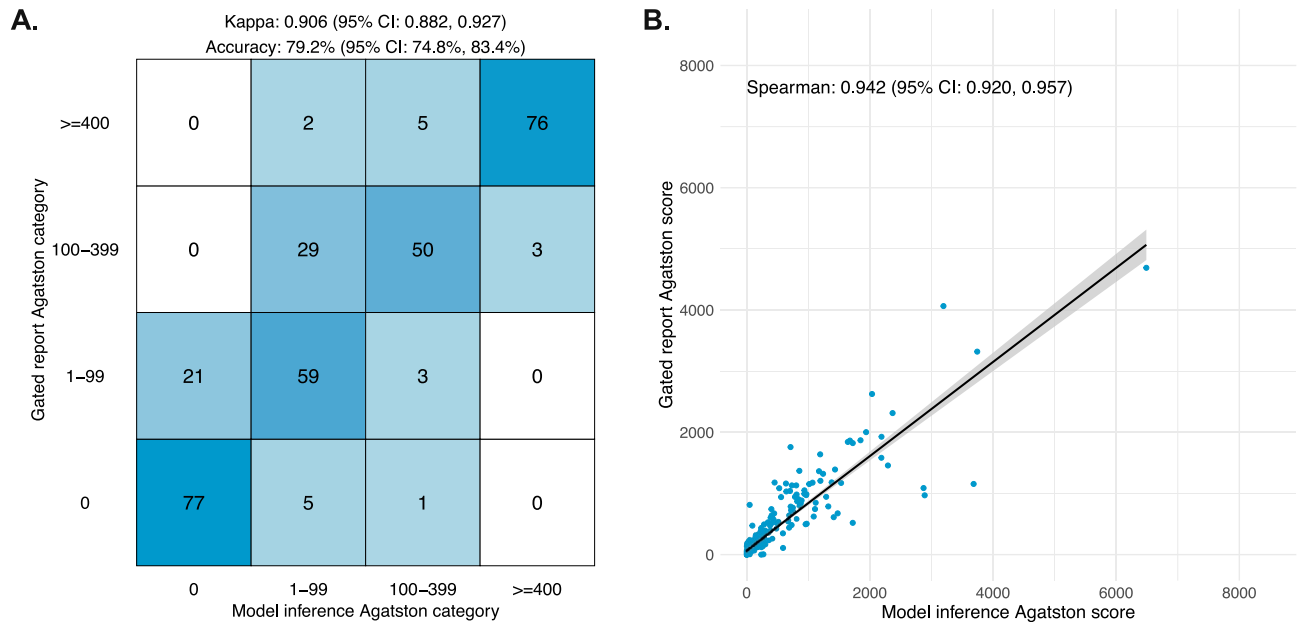


Fig. 4. Comparison of AI model output on non-gated CT with previously reported Agatston categories (**A**) and scores (**B**) from paired cardiac gated CT.

score correlation overall (i.e., based on the sum of the four vessels) remains high despite the lower correlations for these specific vessels. Similar to the previous paragraph, a possibility moving forward is for an AI model to correct for cardiac motion including by incorporating cardiac motion simulations.

A key limitation of this study is that it is a retrospective study outside of the clinical workflow. This study therefore establishes the accuracy of the model but does not assess its impact on the clinical workflow and ultimately on clinical outcomes. Key future research questions include whether the model decreases the need for separate gated cardiac CT cases, how consistently its outputs are incorporated into the radiology reports for non-gated cardiac CT cases and whether those outputs lead to changes in preventative strategies for coronary artery disease.

In conclusion, this study demonstrates that this AI model accurately quantifies coronary artery calcium on non-gated CTs, with strong agreement when compared to both ground truth radiologist interpretations and paired gated CT Agatston scores and categories. Its use in routine non-gated CT scans presents a significant possibility for opportunistic screening, enabling earlier detection of coronary artery disease without the need for specialized cardiac imaging. It could therefore broaden access to screening, facilitate timely intervention and ultimately improve cardiovascular health outcomes.

Data availability

The data used in this study consist of medical imaging data from patient records that are not shared publicly. Access to the data may be considered upon reasonable request, and subject to appropriate institutional and ethical approvals. The corresponding author (Suzannah E McKinney) may be contacted to request data from the study.

Received: 21 March 2025; Accepted: 6 October 2025

Published online: 11 November 2025

References

1. Tsao C. W. *et al.* Heart Disease and Stroke Statistics—2023 Update: A report from the American Heart Association. *Circulation* **147**, (2023).
2. Greenland, P. *et al.* ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: Executive summary. *Circulation* **122**, 2748–2764 (2010).
3. Agatston A. S. *et al.* Quantification of coronary artery calcium using ultrafast computed tomography. *JACC*. **15** (1990).
4. Detrano, R. *et al.* Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N. Engl. J. Med.* **358**, 1336–1345 (2008).
5. Greenland, P. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. *JAMA* **291**, 210 (2004).
6. Orimoloye O. A. *et al.* Race/Ethnicity and the prognostic implications of coronary artery calcium for all-cause and cardiovascular disease mortality: The coronary artery calcium consortium. *J. Am. Heart Assoc.* **7**, (2018).
7. Raggi, P. *et al.* Coronary artery calcium to predict all-cause mortality in elderly men and women. *J. Am. Coll. Cardiol.* **52**, 17–23 (2008).
8. Budoff, M. J. *et al.* Long-term prognosis associated with coronary calcification. *J. Am. Coll. Cardiol.* **49**, 1860–1870 (2007).
9. Polonsky, T. S. Coronary artery calcium score and risk classification for coronary heart disease prediction. *JAMA* **303**, 1610 (2010).
10. Woodard, P. K. & Kovar, C. The direct costs of coronary CT angiography relative to contrast-enhanced thoracic CT: Time-driven activity-based costing. *J. Cardiovasc. Comput. Tomogr.* **15**, 484 (2021).

11. Grant, J. K. et al. Overcoming barriers to implementation: Improving incidental coronary calcium reporting on non-EKG gated chest CT scans. *J. Clin. Lipidol.* <https://doi.org/10.1016/j.jacl.2024.04.129> (2024).
12. Osborne-Grinter, M., Ali, A. & Williams, M. C. Prevalence and clinical implications of coronary artery calcium scoring on non-gated thoracic computed tomography: A systematic review and meta-analysis. *Eur. Radiol.* **34**, 4459–4474 (2023).
13. Topol, E. J. AI-enabled opportunistic medical scan interpretation. *Lancet.* **403**, 1842 (2024).
14. Harris, P. A. et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
15. Harris, P. A. et al. The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
16. Budoff, M. J. et al. Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: The multi-ethnic study of atherosclerosis (MESA). *Eur. Heart J.* **39**, 2401–2408 (2018).
17. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9351** 234–241 (Springer Verlag, 2015).
18. Madubata, C. et al. AI empowering early detection Of CAD patients for improved cardiac care. *J. Cardiovasc. Comput. Tomogr.* **18**, S20 (2024).
19. Han-soo, L. Coreline Soft win FDA 510(k) clearance for AI-powered coronary artery calcification analysis solution. *Korean Biomed. Rev.* (2024).
20. Kang, H. W. et al. Evaluation of fully automated commercial software for Agatston calcium scoring on non-ECG-gated low-dose chest CT with different slice thickness. *Eur. Radiol.* **33**, 1973–1981 (2022).
21. US Food and Drug Administration. K230223 (iCAC Device). https://www.accessdata.fda.gov/cdrh_docs/pdf23/K230223.pdf
22. Eng, D. et al. Automated coronary calcium scoring using deep learning with multicenter external validation. *NPJ. Digit. Med.* **4**, 88 (2021).

Acknowledgements

The authors thank the broader Mass General Brigham AI and Riverain Technologies teams for their assistance with this project.

Author contributions

SEM led the study design and conduct as clinical lead, and was the primary author of the manuscript. SFM conducted the statistical analyses. JKC contributed to case selection, image quality review and series selection. AG contributed to data analysis and management throughout the project. MAH contributed to study management and documentation. SSH, NMM and BG provided ground truth interpretations. KJD was the senior sponsor of the study. MK provided radiological clinical guidance around study design. BCB was the co-Principal Investigator of the study. JMH was the co-Principal Investigator of the study and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by Riverain Technologies. Riverain Technologies were involved in the design of the study; preparation, review and approval of the manuscript; and decision to submit the manuscript for publication. Riverain Technologies were not involved in the collection, management, analysis, and interpretation of the data, which were performed by the named authors who were all employees at Mass General Brigham and its affiliated hospitals. The authors have no other competing interests.

Declarations

Competing interests

All authors declare no financial or non-financial competing interests.

Disclosures

Authors are employees of Mass General Brigham and/or Massachusetts General Hospital, which had received institutional funding from Riverain Technologies for the study.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23183-9>.

Correspondence and requests for materials should be addressed to S.E.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025