

1 **Title Page**

2 Title: Evaluation of an artificial intelligence model for opportunistic calculation of Agatston  
3 score on non-gated computed tomography of the chest

4  
5 Authors:

6 Suzannah E McKinney, MBChB<sup>1,2</sup>

7 Sarah F Mercaldo, PhD<sup>1,2,3</sup>

8 John K Chin, MD<sup>1</sup>

9 Ankita Ghatak, MSc<sup>1</sup>

10 Madeleine A Halle, BSc<sup>1</sup>

11 Sandeep S Hedgire, MD<sup>2,3</sup>

12 Nandini M Meyersohn, MD<sup>2,3</sup>

13 Brian Ghoshhajra, MD, MBA<sup>2,3</sup>

14 Keith J Dreyer, DO PhD<sup>1, 2,3</sup>

15 Mannudeep Kalra, MD<sup>2,3</sup>

16 Bernardo C Bizzo, MD PhD\*<sup>1,2,3</sup>

17 James M Hillis, MBBS DPhil\*<sup>1,3,4</sup>

18

19

20 \* These authors contributed equally to the work and share last authorship.

21

22 Author Affiliations:

23 <sup>1</sup>Mass General Brigham AI, Boston, MA, USA

24 <sup>2</sup>Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

25 <sup>3</sup>Harvard Medical School, Boston, MA, USA

26 <sup>4</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

27

28

29 Corresponding author:

30 Suzannah E McKinney,

31 Mass General Brigham AI

32 399 Revolution Dr, MA 02145

33 semckinney@mgh.harvard.edu

34 857 233 6279

35

36 Date of revision: November 20 2024

37

## 38 **Abstract**

39 **Importance:** The Agatston score is a measure of cardiovascular disease traditionally calculated  
40 on cardiac gated computed tomography (CT) of the chest. Cardiac gated CT is resource-  
41 intensive, can be hard to access, and involves extra radiation exposure. Artificial intelligence  
42 (AI) can be used to opportunistically calculate Agatston score on non-gated CTs performed for  
43 other indications.

44 **Objective:** This study compared the accuracy of an AI model (Riverain Technologies ClearRead  
45 CT CAC) at calculating Agatston scores on non-gated CTs to both consensus radiologist  
46 interpretations on the same CTs and Agatston scores from paired cardiac gated CTs.

47 **Design:** A retrospective standalone performance assessment was conducted on a dataset of non-  
48 contrast CT chest cases acquired between January 2022 and December 2023.

49 **Setting:** The study was conducted at five hospitals in the United States.

50 **Participants:** The cohort included non-gated CTs from 491 patients. It was enriched to ensure a  
51 representation of disease severity by selecting approximately two-thirds of patients using the  
52 originally reported Agatston score on a paired cardiac gated CT within the study timeframe.

53 **Main Outcome(s) and Measure(s):** The study compared the agreement of Agatston categories  
54 (0, 1-99, 100-399 and  $\geq 400$ ) between the AI model and ground truth radiologists or original  
55 radiology reports using the quadratic weighted Kappa coefficient.

56 **Exposure(s):** Each non-gated CT case was interpreted independently by three radiologists to  
57 establish consensus interpretations. Each CT was then interpreted by the AI model. The Agatston  
58 scores for paired cardiac gated CTs were obtained from original radiology reports.

59 **Results:** The agreement between the AI model and ground truth radiologists was 0.959 (95% CI:  
60 0.943-0.975). This result was broadly consistent across sex, age group, race, ethnicity and CT  
61 scanner manufacturer subgroups. The agreement between the AI model and paired cardiac gated  
62 CT was 0.906 (95% CI: 0.882-0.927).

63 **Conclusions and Relevance:** The assessed AI model accurately calculated Agatston scores on  
64 non-gated CTs and produced similar scores to paired cardiac gated CTs. Its use could broaden  
65 screening for atherosclerotic cardiovascular disease, enabling opportunistic screening on CTs  
66 captured for other indications.

## 67 **Introduction**

68 The prevalence of cardiovascular disease (CVD) in adults in the United States is 48.6% (127.9  
69 million people).<sup>1</sup> Heart disease and stroke claim more lives each year than cancer and chronic  
70 lower respiratory disease combined. Moreover, CVD has huge economic impacts with the annual  
71 direct and indirect costs estimated to be \$407.3 billion in the United States in 2018 to 2019.  
72 Morbidity and mortality can be significantly reduced through primary and secondary prevention  
73 strategies, guided by individualized risk assessments. This approach is particularly important when  
74 patients remain asymptomatic,<sup>2</sup> although a challenge for many asymptomatic patients is that they  
75 may not have undergone as extensive diagnostic work-up to establish their risk assessment.

76  
77 Computed tomography (CT) coronary artery calcium (CAC) scoring provides an effective,  
78 noninvasive method for predicting CVD. It involves the measurement of Agatston score, which  
79 is calculated by multiplying the area of calcifications by a factor derived from their highest  
80 attenuation value and then obtaining the sum of these products.<sup>3</sup> An Agatston score can then be  
81 categorized into four groups for disease stratification. This technique has demonstrated its ability  
82 to reliably assess risk across various populations and provides particular value for  
83 prognostication of patients classified as intermediate risk by traditional risk models.<sup>4,5</sup> One  
84 research group followed a total of 42,224 patients of diverse races and ethnicities for a median of  
85 11.7 years; they found that CAC severity based on Agatston category (0, 1-99, 100-399 and  
86  $\geq 400$ ) correlated with risk of all-cause and CVD mortality in all studied race/ethnicity groups.  
87 Further, CAC severity was correlated with all cause and CVD mortality in groups who may be  
88 poorly represented in patient cohorts informing the standard risk prediction models otherwise  
89 recommended by professional bodies.<sup>6</sup> Other large cohort studies have further established

90 Agatston scoring as a valuable tool to assess future CVD risk across different age and ethnicity  
91 subgroups.<sup>7-9</sup>

92

93 A key challenge with CT CAC measurement is that it has traditionally involved gating the  
94 acquisition of the CT with cardiac rhythm to reduce motion artifact. This approach demands  
95 specialized training of radiology technicians and requires patients to undergo a CT for this  
96 limited purpose with the associated radiation exposure.<sup>10</sup> However, there is the opportunity to  
97 use routine, non-gated chest CTs that are obtained for non-cardiac indications, which show CAC  
98 but whose presence is not routinely reported.<sup>11</sup> A 2024 meta-analysis of 108 studies concluded  
99 that identification of incidental CAC on non-gated thoracic CT is a useful technique given the  
100 occurrence of CAC on non-gated scans was both related to the presence of traditional  
101 cardiovascular risk factors (except for smoking status and body mass index) and was predictive  
102 of the development of subsequent cardiovascular events and all-cause mortality.<sup>12</sup>

103

104 A further challenge is the time required by radiology technicians or radiologists to create manual  
105 segmentations of calcified regions to calculate an Agatston score. Artificial Intelligence (AI) has  
106 been used opportunistically in medical imaging elsewhere<sup>13</sup> and the use of AI to automate  
107 manual segmentation has been proposed to enable opportunistic screening on non-gated chest  
108 CT. This study assessed the performance of an AI device (Riverain Technologies ClearRead | CT  
109 CAC), which was designed to estimate the Agatston score by accurately segmenting CAC on  
110 non-contrast, non-gated chest CT.

111

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

## 117 **Methods**

### 118 **Study design**

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

127

### 128 **Case selection**

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

137

138 For the first method, a list of paired cases was created, where patients had received both a gated  
139 CT chest and non-gated CT chest within the study time period (average duration ( $\pm$  standard

140 deviation) between the two studies of  $219.0 \pm 175.1$  days). The reason for this pairing was that  
141 the Agatston score is typically only reported as part of a gated chest CT. The paired approach  
142 therefore enabled enrichment of the dataset to ensure the presence of cases across the spectrum  
143 of CAC severity; albeit the non-gated chest CT cases were ultimately used in the study. These  
144 cases were ordered prospectively by date and then selected consecutively for each of the four  
145 Agatston score categories, in a balanced manner across the hospitals, until there were 83 cases in  
146 each category (i.e., 332 cases total).

147

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

152

### 153 **Image quality review and series selection**

154 All cases were deidentified and then underwent an image quality review by an American Board  
155 of Radiology (ABR)-certified radiologist. Cases were excluded if they contained evidence of  
156 prior cardiac surgery, presence of intravenous or oral contrast, presence of para-cardiac metal  
157 artifact, presence of significant motion artifact that precluded assessment of CAC or presence of  
158 significantly altered thoracic anatomy.

159

160 At the same time as the image quality review, the radiologist selected the series to be given to the  
161 model. To facilitate a range of series slice thicknesses, the radiologist selected up to two soft or  
162 standard kernel axial series as available: one that was  $>0.5\text{mm}$  and  $\leq 1.5\text{mm}$ , and one that was



163 >1.5mm and  $\leq 3$ mm. When a case had both series available and they both passed the DICOM  
164 metadata review as described below, a single series was subsequently selected using  
165 randomization that ensured an even balance between these two slice thickness ranges. The  
166 ground truth radiologists and the AI model only received this single selected series. The image  
167 quality review and series selection process was performed using the FDA-cleared eUnity image  
168 visualization software (Version 6 or higher) and an internal web-based annotation system that  
169 utilized the REDCap electronic data capture tools hosted at MGB.<sup>14,15</sup>

170

### 171 **DICOM metadata review**

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

177

### 178 **Ground truth interpretations**

179 The ground truth interpretations were performed by three ABR-certified radiologists. They  
180 independently segmented the regions of CAC using the TeraRecon iNtuition software (version  
181 4.7.0.22-111). As part of this process, they identified the region of CAC as being in the right  
182 coronary artery, left anterior descending artery, left circumflex artery and left main coronary  
183 artery. The software then calculated the volume of these regions and multiplied them by the  
184 Agatston factor to achieve individual Agatston scores for each vessel as well as an overall  
185 Agatston score.<sup>3</sup> The continuous Agatston scores were also categorized using established

186 categories: 0, 1-99, 100-399 and  $\geq 400$ .<sup>16</sup> We note that Agatston score is traditionally calculated  
187 based on the area on 2.5mm or 3mm slice thickness series from cardiac gated CT; the volume  
188 calculated here was scaled to match 3mm series. In addition, for ease of explanation, we have  
189 referred to “Agatston score” for any measurements of CAC that incorporate a volume or area  
190 multiplied by the Agatston factor while recognizing that its strictest definition only applies to the  
191 overall score obtained on cardiac gated CT.

192

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

200

## 201 **Model inference**

202 The evaluated AI model was version 1.1.1.37 of the Riverain Technologies ClearRead | CT CAC  
203 device. In brief, it consisted of a U-Net style encoder-decoder architecture that operates across  
204 the slices labeling calcified components into one of several categories, including aortic wall and  
205 valve calcifications.<sup>17</sup> The model segments CAC, normalizes for series abnormalities and  
206 calculates an Agatston score. It was trained on approximately 2,000 chest CTs with varied  
207 protocols and acquisition parameters. The model was installed at MGB for use in this study and  
208 received only the relevant series from the Digital Imaging and Communications in Medicine

209 (DICOM)-formatted non-gated CT chest cases. It outputted a continuous score for each vessel  
210 (right coronary artery, left anterior descending artery, left circumflex artery and left main  
211 coronary artery), an overall continuous Agatston score, and an overall Agatston category. The  
212 model also provided an intermediate output of the segmentation masks for any identified areas of  
213 CAC.

214

### 215 **Statistical analysis**

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

223

224 The predefined primary endpoint was the accuracy of the model in calculating the Agatston  
225 category as defined by the quadratic weighted Kappa coefficient when compared with the  
226 consensus ground truth radiologist Agatston category. The percentage accuracy (i.e., proportion  
227 of cases with matching categories) was also calculated. The sample size had been determined  
228 based on powering to ensure the lower bound of the 95% CI of the Kappa coefficient was at least  
229 0.85 using preliminary results.

230

231 The predefined analyses for comparing the model to the ground truth radiologist interpretations  
232 included the accuracy of the overall and component vessel continuous Agatston scores as defined  
233 by the Spearman correlation coefficient. There were also predefined subgroup analyses of the  
234 overall Agatston categories and scores for sex, age, race, ethnicity, CT scanner manufacturer and  
235 radiation dose protocol. Analyses were performed for comparing the paired gated CT overall  
236 Agatston categories and scores to both the model outputs and ground truth radiologist  
237 interpretations. These analyses continued to use the quadratic weighted Kappa coefficient and  
238 percentage accuracy for Agatston categories, and Spearman correlation coefficient for the  
239 Agatston scores.

240

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

## 254 **Results**

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

260

261 The agreement of the overall Agatston category based on quadratic weighted kappa between the  
262 model and consensus ground truth interpretations was 0.959 (95% CI: 0.943 to 0.975) with  
263 accuracy of 92.3% (95% CI: 89.9% to 94.6%; Figure 2). The correlation of the overall  
264 continuous Agatston score between the model and the average ground truth radiologist  
265 interpretations as defined by the Spearman correlation coefficient was 0.975 (95% CI: 0.962 to  
266 0.987). The agreement was broadly consistent across sex, age, ethnicity, race, manufacturer and  
267 radiation dose protocol subgroups. All subgroups had a kappa coefficient of at least 0.90 except  
268 for the “Other” race subgroup, which had a kappa coefficient of 0.855 for only 11 cases (Table  
269 S1).

270

271 The correlation of Agatston scores for individual arteries showed a Spearman coefficient of  
272 0.951 (95% CI: 0.925 to 0.971) for the right coronary artery, 0.985 (95% CI: 0.978 to 0.991) for  
273 the left anterior descending artery, 0.867 (95% CI: 0.824 to 0.904) for the left circumflex artery  
274 and 0.791 (95% CI: 0.736 to 0.839) for the left main coronary artery (Figure 3).

275

276 The Dice scores were calculated based on 368 segmentations available from the model and 291,  
277 287 and 288 segmentations from the three ground truth radiologists. The average Dice score  
278 between the model and ground truth radiologists was 0.892 (95% CI: 0.871 to 0.911) and the  
279 average Dice score between ground truth radiologists was 0.910 (95% CI: 0.893 to 0.927). The  
280 average difference between Dice scores was -0.008 (95% CI: -0.024 to 0.007), and there was no  
281 evidence of a difference between the distributions of Dice scores ( $p = 0.12$ ).

282

283 When considering the agreement of the Agatston categories between the model output on the  
284 non-gated CT and the reported values for the paired cardiac gated CT, the kappa coefficient was  
285 0.906 (95% CI: 0.882 to 0.927) with accuracy of 79.2% (95% CI: 74.8% to 83.4%; Figure 4).

286 The Spearman correlation coefficient of the Agatston score was 0.942 (95% CI: 0.920 to 0.957).

287 The comparisons between the ground truth radiologist interpretations on the non-gated CT and  
288 the reported values for the paired cardiac gated CT had kappa coefficient of 0.907 (95% CI:

289 0.883 to 0.930), accuracy of 79.5% (95% CI: 75.2% to 83.6%) and Spearman correlation

290 coefficient of 0.941 (95% CI: 0.920 to 0.957; Figure S1).

## 291 **Discussion**

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

299  
300 These results demonstrate that the model can accurately quantify Agatston scores on non-gated  
301 CTs. They show a potential clinical use in screening for incidental coronary artery disease during  
302 routine CTs. Importantly, these results also appear to generalize across demographic and  
303 technical subgroups, a critical prerequisite for a model to perform consistently across diverse  
304 populations.<sup>6</sup> The economic benefits of cardiac gated CTs versus more invasive measures such as  
305 cardiac angiography have previously been shown, as have the benefits of risk stratification and  
306 modification compared to untreated cardiovascular disease.<sup>10,18</sup> It may be possible to realize even  
307 greater benefits given the higher number of people undergoing routine CTs compared to cardiac  
308 gated CTs. This AI model could therefore truly enable opportunistic screening for atherosclerotic  
309 cardiovascular disease.

310  
311 Validation studies that compare AI model quantification of coronary artery calcium on non-gated  
312 CT to ground truth results from cardiac gated CT have been reported. The FDA-cleared AVIEW  
313 CAC device by Coreline<sup>19</sup> was assessed on a cohort of 567 patients for five categories of

314 Agatston scores (it used the categories 0, 1-10, 11-100, 101-400 and >400); it specifically used  
315 low-dose non-gated CT.<sup>20</sup> This model achieved Cohen's weighted kappa of 0.809 (95% CI:  
316 0.776 to 0.838) with slice thickness 1mm and 0.776 (95% CI: 0.740 to 0.809) with slice  
317 thickness 2.5mm. A separate model, which involved team members from Bunkerhill who have  
318 subsequently received FDA clearance for the iCAC device<sup>21</sup>, achieved kappa  
319 coefficients of 0.802, 0.684, 0.644 and 0.583 on a cohort of 303 cases from four different sites.<sup>22</sup>  
320 The current model compared favorably by achieving a kappa coefficient of 0.906 (95% CI: 0.882  
321 to 0.927) for a similar analysis.

322

323 Although these previous works generate valuable proof points about the validity of using AI  
324 models to calculate Agatston scores from non-gated CT, and therefore the enablement of  
325 opportunistic screening on non-gated CT, our study here goes further to demonstrate the viability  
326 of opportunistic screening scoring on non-gated studies in three important ways. Firstly, we  
327 account for acquisition parameter differences between gated and non-gated scans. As the 2021  
328 paper points out, because non-gated chest CT are performed for non-cardiac indications,  
329 acquisition parameters differ for gated coronary CTs which can affect the accuracy of calcium  
330 quantitation. Our study intentionally includes cases from both low-dose and routine-dose  
331 protocols, as well as varying slice thicknesses, to minimize the impact of acquisition parameter  
332 differences when comparing gated and non-gated scans. Secondly, we show stronger correlation  
333 between the AI model Agatston score output and consensus ground truth than previously seen.  
334 Thirdly, we used as our primary source of ground truth Agatston scores derived from the manual  
335 segmentations of three radiologists on a large non-gated dataset of 491 patients.

336



337 Although agreement was high all around, there was higher agreement in our study [0.959 (95%  
338 CI: 0.943-0.975)] between both the AI model output and consensus ground truth than AI model  
339 and Agatston scores from paired gated studies [0.906 (95% CI: 0.882-0.927)] and between  
340 consensus ground truth and Agatston scores from paired gated studies [0.907 (95% CI: 0.883,  
341 0.930)]. Although CAC develops very slowly over time and time periods of <1 year are unlikely  
342 to lead to disparate changes between studies, there may be some small impact from the temporal  
343 delay between paired studies. Perhaps more importantly, non-gated studies offer a higher spatial  
344 resolution than gated due to a thinner typical slice thickness; it is therefore possible that non-  
345 gated studies detect calcium that is not seen in gated studies. However, non-gated scans may  
346 contain cardiac motion that obfuscates calcium measurement and contain generally more noise,  
347 which may quantification of calcium more challenging.

348

349 At a coronary vessel level, the model showed better performance for the right coronary artery  
350 and left anterior descending artery (Spearman correlation coefficient of 0.951 and 0.985  
351 respectively) than for the left circumflex artery and left main coronary artery (0.867 and 0.791  
352 respectively). A key possible reason is that the left circumflex artery and left main coronary  
353 artery are more susceptible to cardiac motion and therefore more challenging to segment. The  
354 left circumflex artery (LCX) and left main coronary artery (LM) are particularly affected by  
355 respiratory and cardiac motion due to their lateral and anterior positions, with the LCX being  
356 more tortuous and susceptible to motion artifacts. Additionally, the LCX's smaller, more curved  
357 vessels experience greater displacement from diaphragmatic motion, while the LM is influenced  
358 by the significant movement of the heart's base during the cardiac cycle. In addition, the  
359 delineation of the left circumflex artery and the left main coronary artery may not be as clear,

360 which may cause inconsistencies between the ground truth radiologists and the model in  
361 attributing calcium to either of these vessels. This situation may reflect why the Agatston score  
362 correlation overall (i.e., based on the sum of the four vessels) remains high despite the lower  
363 correlations for these specific vessels.

364

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

371

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

379 **Acknowledgements**

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

382

383 **Source of Funding**

384 This study was funded by Riverain Technologies. Riverain Technologies was involved in the  
385 design and conduct of the study; preparation, review, and approval of the manuscript; and  
386 decision to submit the manuscript for publication. Riverain Technologies was not involved in the  
387 collection, management, analysis, and interpretation of the data.

388

389 **Disclosures**

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

392

393 **Supplemental Material**

394 Figure S1, Table S1

## 395 References

- 396 1. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics—2023  
397 Update: A Report From the American Heart Association. *Circulation*. 2023;147(8).  
398 doi:10.1161/CIR.0000000000001123
- 399 2. Greenland P, Alpert JS, Beller GA, et al. 2010 ACCF/AHA Guideline for Assessment of  
400 Cardiovascular Risk in Asymptomatic Adults: Executive Summary. *Circulation*.  
401 2010;122(25):2748-2764. doi:10.1161/CIR.0b013e3182051bab
- 402 3. Agatston AS, Janowitz WR, Hildner FJ, et al. *Quantification of Coronary Artery Calcium*  
403 *Using Ultrafast Computed Tomography*. Vol 15.; 1990.
- 404 4. Detrano R, Guerci AD, Carr JJ, et al. Coronary Calcium as a Predictor of Coronary Events  
405 in Four Racial or Ethnic Groups. *New England Journal of Medicine*. 2008;358(13):1336-  
406 1345. doi:10.1056/NEJMoa072100
- 407 5. Greenland P. Coronary Artery Calcium Score Combined With Framingham Score for Risk  
408 Prediction in Asymptomatic Individuals. *JAMA*. 2004;291(2):210.  
409 doi:10.1001/jama.291.2.210
- 410 6. Orimoloye OA, Budoff MJ, Dardari ZA, et al. Race/Ethnicity and the Prognostic  
411 Implications of Coronary Artery Calcium for All-Cause and Cardiovascular Disease  
412 Mortality: The Coronary Artery Calcium Consortium. *J Am Heart Assoc*. 2018;7(20).  
413 doi:10.1161/JAHA.118.010471
- 414 7. Raggi P, Gongora MC, Gopal A, Callister TQ, Budoff M, Shaw LJ. Coronary Artery  
415 Calcium to Predict All-Cause Mortality in Elderly Men and Women. *J Am Coll Cardiol*.  
416 2008;52(1):17-23. doi:10.1016/j.jacc.2008.04.004
- 417 8. Budoff MJ, Shaw LJ, Liu ST, et al. Long-Term Prognosis Associated With Coronary  
418 Calcification. *J Am Coll Cardiol*. 2007;49(18):1860-1870. doi:10.1016/j.jacc.2006.10.079
- 419 9. Polonsky TS. Coronary Artery Calcium Score and Risk Classification for Coronary Heart  
420 Disease Prediction. *JAMA*. 2010;303(16):1610. doi:10.1001/jama.2010.461
- 421 10. Woodard PK, Kovar C. The Direct Costs of Coronary CT Angiography Relative to  
422 contrast-enhanced thoracic CT: Time-driven activity-based costing. *J Cardiovasc Comput*  
423 *Tomogr*. 2021;15(6):484. doi:10.1016/j.jcct.2021.07.003
- 424 11. Grant JK, Bokhari A, Manoharan A, et al. Overcoming barriers to implementation:  
425 Improving incidental coronary calcium reporting on non-EKG gated chest CT scans. *J*  
426 *Clin Lipidol*. Published online April 2024. doi:10.1016/j.jacl.2024.04.129
- 427 12. Osborne-Grinter M, Ali A, Williams MC. Prevalence and clinical implications of coronary  
428 artery calcium scoring on non-gated thoracic computed tomography: a systematic review  
429 and meta-analysis. *Eur Radiol*. 2023;34(7):4459-4474. doi:10.1007/s00330-023-10439-z
- 430 13. Topol EJ. AI-enabled opportunistic medical scan interpretation. *The Lancet*.  
431 2024;403(10439):1842. doi:10.1016/S0140-6736(24)00924-3
- 432 14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data  
433 capture (REDCap)—A metadata-driven methodology and workflow process for providing  
434 translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.  
435 doi:10.1016/j.jbi.2008.08.010
- 436 15. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international  
437 community of software platform partners. *J Biomed Inform*. 2019;95:103208.  
438 doi:10.1016/j.jbi.2019.103208

- 439 16. Budoff MJ, Young R, Burke G, et al. Ten-year association of coronary artery calcium with  
440 atherosclerotic cardiovascular disease (ASCVD) events: the multi-ethnic study of  
441 atherosclerosis (MESA). *Eur Heart J*. 2018;39(25):2401-2408.  
442 doi:10.1093/eurheartj/ehy217
- 443 17. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image  
444 segmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes*  
445 *in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 9351. Springer Verlag;  
446 2015:234-241. doi:10.1007/978-3-319-24574-4\_28
- 447 18. Madubata C, Lim N, Kimmel Y, et al. AI Empowering Early Detection Of CAD Patients  
448 For Improved Cardiac Care. *J Cardiovasc Comput Tomogr*. 2024;18(4):S20.  
449 doi:10.1016/j.jcct.2024.05.044
- 450 19. Han-soo L. Coreline Soft win FDA 510(k) clearance for AI-powered coronary artery  
451 calcification analysis solution. *Korean Biomedical Review*.
- 452 20. Kang HW, Ahn WJ, Jeong JH, et al. Evaluation of fully automated commercial software  
453 for Agatston calcium scoring on non-ECG-gated low-dose chest CT with different slice  
454 thickness. *Eur Radiol*. 2022;33(3):1973-1981. doi:10.1007/s00330-022-09143-1
- 455 21. FDA. *K230223 Clearance Document*. <https://www.fda.gov/medical->
- 456 22. Eng D, Chute C, Khandwala N, et al. Automated coronary calcium scoring using deep  
457 learning with multicenter external validation. *NPJ Digit Med*. 2021;4(1):88.  
458 doi:10.1038/s41746-021-00460-1  
459

460

461 **Table 1:** Demographic and technical breakdown of CT chest cases.

	<b>Agatston score 0</b>	<b>Agatston score 1-99</b>	<b>Agatston score 100-399</b>	<b>Agatston score ≥400</b>	<b>Overall</b>
<b>Total cases</b>	143	125	84	139	491
<b>Sex</b>					
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%)
<b>Age</b>					
≤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
<b>Race</b>					
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%)
<b>Ethnicity</b>					

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%)
<b>Manufacturer</b>					
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%)
<b>Radiation dose protocol</b>					
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%)
<b>Paired cardiac-gated CT</b>					
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

462  
463

464 **Figure Legends**

465 **Figure 1:** Cohort diagram.

466

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

469

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

473

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

476



477 **Table S1:** Subgroup analyses for demographic and technical variables for the agreement of  
 478 Agatston categories (Kappa and accuracy) and correlation of Agatston score (Spearman  
 479 coefficient) between the AI model output and ground truth radiologists on non-gated CT.

480

<b>Subgroup</b>	<b>N</b>	<b>Kappa (95% CI)</b>	<b>Accuracy (95% CI)</b>	<b>Spearman <math>\rho</math> (95% CI)</b>
<b>Sex</b>				
Female	272	0.957 (0.935, 0.975)	91.5 (88.2, 94.8)	0.965 (0.937, 0.985)
Male	219	0.962 (0.936, 0.982)	93.2 (89.8, 96.6)	0.986 (0.974, 0.993)
<b>Age</b>				
<65 years	251	0.960 (0.938, 0.979)	92.4 (89.1, 95.8)	0.972 (0.949, 0.989)
>65 years	240	0.958 (0.933, 0.978)	92.1 (88.6, 95.5)	0.973 (0.956, 0.987)
<b>Race</b>				
Asian	11	0.927 (0.818, 1.000)	81.8 (58.7, 100.0)	0.895 (0.561, 1.000)
Black or African American	7	1.000 (1.000, 1.000)	100.0 (100.0, 100.0)	1.000 (1.000, 1.000)
White	448	0.961 (0.943, 0.976)	92.9 (90.5, 95.2)	0.975 (0.961, 0.987)
Other	11	0.855 (0.745, 0.964)	63.6 (35.2, 92.9)	0.930 (0.646, 1.000)
Declined	12	1.000 (1.000, 1.000)	100.0 (100.0, 100.0)	0.993 (0.925, 1.000)
Unavailable	2	1.000 (1.000, 1.000)	100.0 (100.0, 100.0)	1.000 (1.000, 1.000)
<b>Ethnicity</b>				
Hispanic	12	0.933 (0.833, 1.000)	83.3 (62.0, 100.0)	0.964 (0.826, 1.000)
Not Hispanic	452	0.958 (0.941, 0.973)	92.3 (89.9, 94.8)	0.974 (0.960, 0.986)
Prefer not to say/Declined	7	1.000 (1.000, 1.000)	100.0 (100.0, 100.0)	1.000 (1.000, 1.000)

Unavailable	20	0.980 (0.940, 1.000)	95.0 (85.4, 100.0)	0.994 (0.959, 1.000)
<b>Manufacturer</b>				
GE Medical Systems	216	0.969 (0.944, 0.987)	94.9 (91.9, 97.9)	0.984 (0.969, 0.995)
Philips	20	0.980 (0.940, 1.000)	95.0 (85.2, 100.0)	0.991 (0.958, 1.000)
Siemens	214	0.948 (0.920, 0.970)	89.7 (85.7, 93.9)	0.962 (0.931, 0.983)
Toshiba	41	0.961 (0.922, 0.990)	90.2 (80.9, 99.1)	0.979 (0.934, 0.998)
<b>Radiation dose</b>				
Low dose	269	0.967 (0.951, 0.982)	92.9 (89.8, 96.0)	0.977 (0.959, 0.990)
Routine dose	222	0.950 (0.921, 0.973)	91.4 (87.7, 95.1)	0.972 (0.948, 0.989)

481