



Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT

Cheng Xu¹, Heng Guo², Minfeng Xu², Miao Duan³, Ming Wang¹, Peijun Liu¹, Xinyi Luo^{4,5}, Zhengyu Jin¹, Hui Liu^{4,5,6}, Yining Wang¹

¹Department of Radiology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; ²Alibaba Group, Hangzhou, China; ³Department of Radiology, Shunyi Hospital, Beijing Traditional Chinese Medicine Hospital, Beijing, China; ⁴Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China; ⁵School of Medicine, South China University of Technology, Guangzhou, China; ⁶The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China

Contributions: (I) Conception and design: Y Wang; (II) Administrative support: Z Jin, Y Wang, M Xu, H Liu; (III) Provision of study materials or patients: Z Jin, H Liu, Y Wang; (IV) Collection and assembly of data: M Duan, C Xu, H Guo, P Liu, X Luo; (V) Data analysis and interpretation: C Xu, H Guo, M Xu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Prof. Yining Wang. Department of Radiology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. Email: wangyining@pumch.cn.

Background: The aim of this study was to investigate the reliability and accuracy of automatic coronary artery calcium (CAC) scoring and risk classification in non-gated, non-contrast chest computed tomography (CT) of different slice thicknesses using a deep learning algorithm.

Methods: This retrospective study was performed at 2 tertiary hospitals. Paired, dedicated calcium-scoring CT scans and non-gated, non-contrast chest CT scans taken within a month from the same patients were included. Chest CT images were grouped according to the slice thickness (group A: 1 mm; group B: 3 mm). For internal scans, the CAC score manually measured on dedicated calcium scoring CT images was used as the gold standard. The deep learning algorithm for group A was trained using 150 chest CT scans and tested using 144 scans, and that for group B was trained using 170 chest CT scans and tested using 144 scans. The intraclass correlation coefficient (ICC) was used to evaluate the correlation between the algorithm and the gold standard. Agreement between the deep learning algorithm, the manual results on chest CT, and the gold standard was determined by Bland-Altman analysis. Cardiac risk categories were compared. External validation was performed on 334 paired scans from a different organization.

Results: A total of 608 internal paired scans (1 mm: 294; 3 mm: 314) of 406 individuals and 334 external paired scans (1 mm: 117; 3 mm: 117) of 117 individuals were included in the analysis. The ICCs between the deep learning algorithm and the gold standard were excellent in both group A (0.90; 95% CI: 0.85–0.93) and group B (0.94; 95% CI: 0.92–0.96). The Bland-Altman plots showed good agreement in both groups. For the cardiovascular risk category, the deep learning algorithm accurately classified 71% of cases in group A and 81% of cases in group B. The Kappa values for risk classification were 0.72 in group A and 0.82 in group B. External validation yielded equally good results.

Conclusions: The automatic calculation of CAC score and cardiovascular risk stratification on non-gated chest CT using a deep learning algorithm was reliable and accurate on both 1 and 3 mm scans. Chest CT with a slice thickness of 3 mm was slightly more accurate in CAC detection and risk classification.

Keywords: Coronary artery disease; coronary artery calcium (CAC) score; deep learning; chest computed tomography (CT); atherosclerosis

Submitted Oct 12, 2021. Accepted for publication Jan 24, 2022.

doi: 10.21037/qims-21-1017

View this article at: <https://dx.doi.org/10.21037/qims-21-1017>

Introduction

Cardiovascular disease (CVD) is one of the leading causes of death worldwide. Coronary artery calcium (CAC) scoring is a well-established approach to predicting adverse cardiovascular events (1) and guiding treatment decisions (2,3). Typically, a CAC score is obtained from electrocardiogram (ECG)-gated, non-contrast-enhanced cardiac computed tomography (CT), and a radiologist needs to identify dense calcification on CT images, a process that is time-consuming. In addition, non-contrast chest CT is widely used in clinical practice, especially in lung cancer screening, which shares some risk factors with CVD. However, quantification of the calcification score is often ignored on chest CT. CAC quantifications based on chest CT can provide additional information for clinicians screening for lung disease, thereby improving the efficiency of clinical diagnosis without additional cost and avoiding additional radiation doses. Automatic CAC scoring has attracted increased attention in recent years due to the rapid development of technologies in the deep learning community. The results of CAC scoring can be presented automatically while evaluating chest CT images, requiring no manual operation or review by clinicians.

Strategies previously reported in the literature concerning automatic CAC scoring can be broadly divided into two categories: (I) CAC is first identified and thereafter quantified, similar to a clinical workflow; (II) CAC scoring is directly regressed with a large training dataset, thus eliminating the intermediate identification process. In the first category, the automatic scoring methods typically rely on segmenting or roughly localizing the anatomical structures (4-10), such as the heart, to obtain a region of interest (ROI). For example, Wolterink *et al.* (5) and Lessmann *et al.* (6) both used a dedicated convolutional neural network (ConvNet) to localize the heart with a bounding box. Wolterink *et al.* (9) identified candidate calcifications that could not be automatically labeled with high certainty and optionally presented these to an expert for review. Considering the high computational cost of

the aforementioned automatic methods, particularly when applied to large datasets, some researchers have focused on exploring direct quantification methods for CAC (11-14). Cano-Espinosa *et al.* (11) trained a 3D deep ConvNet to regress the Agatston score within a pre-segmented region of the heart, and yielded a high correlation with manual measurements ($r=0.93$). More recently, de Vos *et al.* (13) employed two ConvNets, one for atlas registration and one for CAC regression, and demonstrated a computationally efficient solution for both cardiac and chest CT, with an intraclass correlation coefficient (ICC) between the predicted and manual calcium scores of 0.98. Despite its efficiency, the direct regression method also has its disadvantages. For example, when regions that contribute to the calcium score are expected to be investigated, a decision feedback process must be executed to obtain a visual attention heatmap. However, in most cases, this heatmap cannot depict lesions with pixel-level accuracy. Additionally, the reference standard for most of the aforementioned studies was a manual result on chest CT, which differs from that measured on a dedicated calcium scoring CT of the same patient, as motion artifacts are more likely to occur on chest CT images due to the lack of ECG control (15). Recently, several studies investigated automatic deep learning algorithms using a dedicated calcium scoring CT as a reference, but the sample sizes were relatively small (16), the interval between the paired scans was long (17), and the influence of slice thickness on the algorithm was not compared.

Therefore, we proposed and validated a two-stage segmentation deep learning algorithm and performed this study to investigate the reliability and accuracy of the automatic CAC scoring and risk classification on non-gated, non-contrast chest CT scans of different slice thicknesses using a deep learning algorithm.

We present the following article in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-21-1017/rc>).

Methods

Patient population

The scans of patients who had undergone both a dedicated calcium scoring CT and a non-gated chest CT within 1 month between October 2016 and July 2019 in our hospital were retrospectively included as internal paired scans, whereas external paired scans were collected from January 2021 to April 2021 in another hospital. The pairing criteria involved matching a non-gated, non-contrast chest CT scan with a dedicated gated, non-contrast CT scan from the same patient using a dedicated CAC-scoring protocol. Patients with severe motion artifacts or a history of stent, pacemaker, artificial valve implantation, or coronary bypass graft surgery were excluded.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Institutional Ethics Committee of the Peking Union Medical College Hospital. Written informed consent was waived due to the retrospective nature of the study.

Cardiac and chest CT protocol

Non-gated chest CT scans were performed using multidetector CT scanners from Siemens AG (Somatom Definition Flash or Somatom Force; Siemens, Forchheim, Germany), General Electric (Discovery CT750 HD; Milwaukee, WI, USA), Philips (IQon Spectral CT, Cleveland, OH, USA), and Toshiba (Aquilion 64; Tokyo, Japan), with a tube voltage of 120 kVp and automatically adjusted tube current. The slice thickness varied from 1 to 5 mm and the slice increments varied from 1 to 3 mm. The scans ranged from the thoracic inlet to the adrenal glands with breath-hold instruction. The chest CT images were grouped according to the slice thickness (group A: 1 mm; group B: 3 mm).

Dedicated calcium scoring CT was performed on 3 different systems with prospective ECG-triggered cardiac CT (IQon Spectral CT, Philips Healthcare; Somatom Definition Flash or Somatom Force, Siemens). The slice thicknesses were 1.5 (IQon Spectral CT) and 3 mm (Somatom Definition Flash or Somatom Force), with a tube voltage of 120 kVp and automatically adjusted tube current. Acquisitions included the entire heart, from the carina to the level of the diaphragm. Intravenous contrast agents were not used in any of the scans.

Calcification score assessment

The CAC scores from the dedicated calcium scoring CT were calculated using commercial software (Syngo.via VB10; Siemens Healthcare), and board-certified radiologists in our hospital with relevant experience performed the CAC score measurements for the clinical reports.

For chest CT, the CAC scores were calculated both manually by a radiologist with 5 years of experience (blinded to the results on dedicated calcium scoring CT) and automatically using a deep learning algorithm. The threshold for calcification detection was set at 130 Hounsfield units (HU) with an area $\geq 1 \text{ mm}^3$. The density score was determined based on the maximal attenuation of lesions, as follows: score 1, 130–199 HU; score 2, 200–299 HU; score 3, 300–399 HU; score 4, ≥ 400 HU. The calcification score for each lesion was calculated by summing the multiplication of the density score and slice area of all slices (18). The total score for each patient was calculated automatically by summing the scores of all the lesions. CAC scores of 0, 1 to 100, 101 to 300, and more than 300 represented a very low risk, mildly increased risk, moderately increased risk, and severely increased risk, respectively (19).

Deep learning algorithm

In this work, a widely used U-Net (20,21) was adopted to segment the heart region into ROIs in the first stage. To efficiently prepare the training data, only CT scans from the training set of group B (slice thickness of 3 mm) were annotated at the pixel level. Specifically, because aortic calcified plaques are difficult to distinguish from CAC, the aortic region and remaining regions inside the heart were marked as 2 different foreground classes during annotation. During the whole inference pipeline, false-positive calcium lesions located in the aortic region were identified and removed. The input data were reshaped uniformly, with respect to both the training and inference of heart segmentation. In our experiments, data were shaped to 64, 256, and 256 pixels in depth, height, and width, respectively. Data with slice thicknesses of 1 mm were also downsampled to this shape to perform heart segmentation.

A semantic segmentation network comprising autofocus layers (22) was employed to identify CAC in the second stage. Using autofocus layers, more powerful features could be effectively extracted because of the adaptive receptive

Table 1 Detailed baseline characteristics of the study population

| Characteristics | Internal cohort (n=406) | External cohort (n=117) |
|---------------------------|----------------------------|----------------------------|
| Age (year) | 61.8±12.1 | 59.8±12.6 |
| Male | 217 (53.3%) | 74 (63.2%) |
| BMI (kg/m ²) | 26.8±4.2 | 25.8±3.5 |
| Diabetes | 78 (19.2%) | 21 (17.9%) |
| Hypertension | 183 (45.1%) | 40 (34.2%) |
| Hypercholesterolemia | 80 (19.6%) | 17 (14.5%) |
| Smoking | 119 (29.3%) | 27 (23.1%) |
| CVD family history | 32 (7.9%) | 1 (1.0%) |
| Median CAC score (25–75%) | 45.6 (1.2–211.6) | 1.4 (0–60.7) |

BMI, body mass index; CVD, cardiovascular disease; CAC, coronary artery calcium.

Table 2 CAC score categories of the test and validation cohorts

| Category | Internal cohort | | External cohort | |
|----------|-----------------|-----------------|-----------------|-----------------|
| | 1 mm (n=144) | 3 mm (n=144) | 1 mm (n=117) | 3 mm (n=117) |
| I | 52 | 49 | 54 | 54 |
| II | 50 | 49 | 39 | 39 |
| III | 21 | 23 | 9 | 9 |
| IV | 21 | 23 | 15 | 15 |

CAC, coronary artery calcium.

fields. 3D chest CTs with calcified plaques were manually annotated pixel-by-pixel and sampled as fixed-size patches for network training. In our experiments, the sampling size was 16, 64, and 64 pixels in depth, height, and width, respectively. During the inference phase, candidate plaques chosen by a threshold of 130 HU inside the ROI were also sampled as fixed-size patches, and then the collected patches were sequentially predicted by the autofocus network, with the overlap between segmented patches determined by a majority voting strategy. In order to test the generalizability of the model, we used a relatively large number of testing datasets. The datasets for training and testing were composed of 150 and 144 CT scans, respectively, in group A. The datasets for training and testing comprised 170 and 144 CT scans, respectively, in group B.

For data preprocessing, the 12-bit CT intensity range was clipped to a range (−125, 225) HU to obtain a

better contrast for heart and calcium lesions. Each image was then normalized to a floating-point number in the 0–1 range. During the training stage, random cropping and random mirroring in the axial plane were adopted for data augmentation. The networks were trained using an Adam optimizer (23) with an initial learning rate of 0.0002, and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.9$. In the total 200 epochs, the learning rate was linearly reduced to 0 during the second 100 epochs.

Statistical analysis

Statistical analyses were performed using SPSS 20.0 (SPSS Inc., Chicago, IL, USA) and MedCalc Statistical Software version 15.2.2 (MedCalc Software bvba, Ostend, Belgium). Continuous variables were expressed as the mean values \pm standard deviation or medians (interquartile range). Categorical variables were expressed in terms of the frequency and composition ratio (%). The Kolmogorov-Smirnov test was used to assess the normality of the quantitative data. The ICC was used to evaluate the correlation between the algorithm and the gold standard. The agreement between the deep learning algorithm, the manual results on chest CT, and the gold standard was determined by Bland-Altman analysis. The Chi-squared test was used to evaluate the ability of the deep learning algorithm to detect calcification. Linearly weighted Kappa values were used to assess the concordance of the CAC score ranking of the chest CT and the dedicated calcium scoring CT. A value of $P < 0.05$ was considered statistically significant.

Results

Patient population

Based on the predefined exclusion criteria, 406 internal individuals (217 men and 189 women) and 117 external individuals (74 men and 43 women) were included in the analysis. *Table 1* shows the detailed baseline characteristics of the study population and *Table 2* shows the CAC score categories of the test and validation cohorts. *Figure 1* shows the flowchart of patient enrollment and study design.

Reliability and accuracy of the deep learning algorithm

The reliability of the deep learning algorithm was defined as the difference between manual and automatic CAC scores on chest CT. The accuracy of the deep learning algorithm

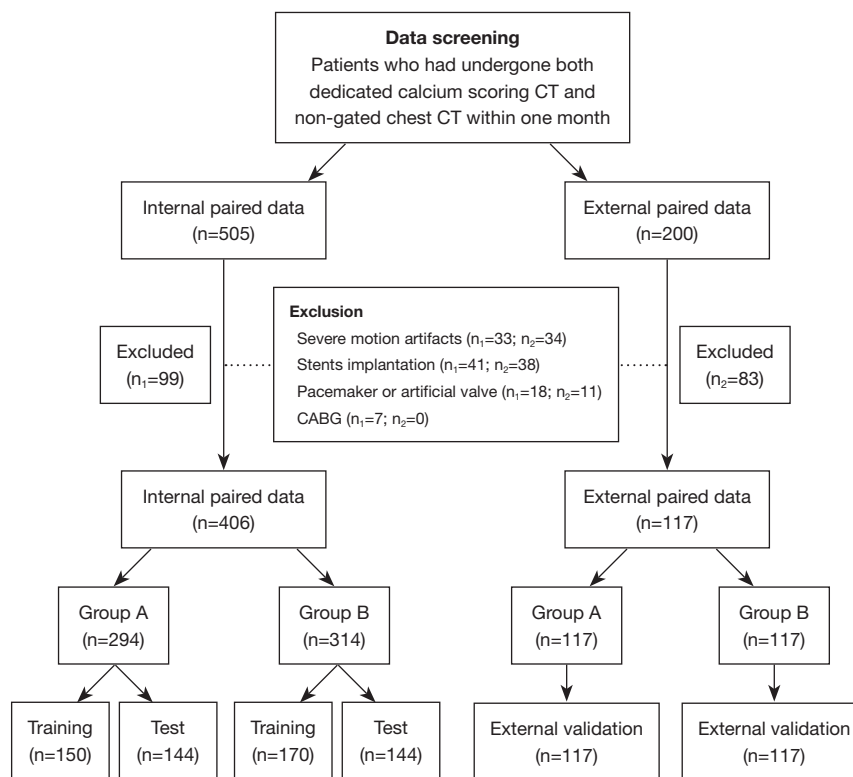


Figure 1 Flowchart of patient enrollment and study design. Chest CT images were grouped according to the slice thickness (group A: 1 mm; group B: 3 mm). CT, computed tomography; CABG, coronary artery bypass grafting.

was defined as the difference between the automatic CAC score on chest CT and the gold standard. The systemic difference was defined as the difference between the manual results on chest CT and the gold standard, which inevitably affected the accuracy of the deep learning algorithm. The reliabilities of the algorithm were good on both the 1 and 3 mm scans. The mean difference (with 95% limits of agreement) between the algorithm on chest CT and the gold standard was slightly better on the 3 mm scans $[-10$ (-296.7 to 276.7) *vs.* -55 (-464.9 to 354.9)]. There were systematic differences in both the 1 and 3 mm thickness images. *Figure 2* shows the detailed Bland-Altman plots. Excellent ICCs between the chest CT and the gold standard were obtained while quantifying the CAC scores for group A (ICC =0.90; 95% CI: 0.85–0.93) and group B (ICC =0.94; 95% CI: 0.92–0.96). *Figure 3* shows several example cases.

CAC detection

Coronary calcifications were detected in 63.8% of cases in group A and 65.9% of cases in group B according to the dedicated calcium scoring CT. Compared with the gold

standard, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy of the automatic algorithm for detecting CAC >0 were, respectively, 90%, 83%, 90%, 83%, and 88% in group A and 99%, 86%, 93%, 98%, and 94% in group B. Nine false-positive cases were identified in group A while 7 cases were identified in group B, and the cause was misidentification of valve calcification. Nine false-negative cases were identified in group A while 1 case was identified in group B, and this situation could be attributed to cardiac motion.

CAC category

Regarding the CAC categories, the agreement between the automatic and gold standards was quite good. The Kappa value was 0.72 (95% CI: 0.64–0.80) in group A and 0.82 (95% CI: 0.76–0.89) in group B. The deep learning algorithm accurately classified 71% (group A) and 81% (group B) of the data. Regarding group B, 22 cases (15%) were reclassified to higher categories and 6 cases (4%) were reclassified to lower categories. *Figure 4* shows the detailed classification confusion matrices.

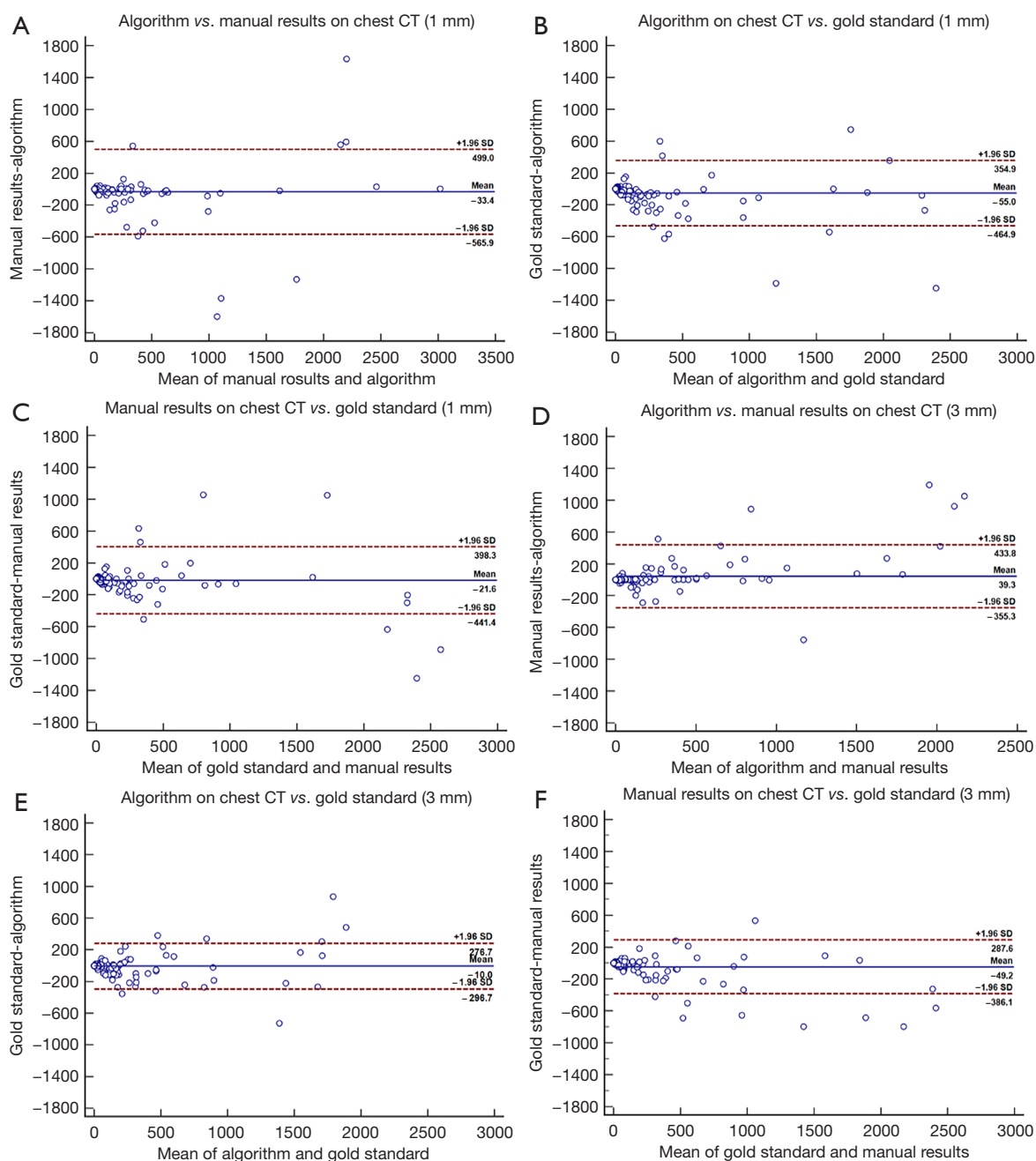


Figure 2 Bland-Altman plots. The Bland-Altman plots show agreement between the deep learning algorithm and the manual results on chest CT scans of 1 mm slice thickness (A) and 3 mm slice thickness (D). The gold standard (manual results on dedicated calcium scoring CT) and the deep learning algorithm on chest CT scans of 1 mm slice thickness (B) and 3 mm slice thickness (E). The gold standard and the manual results on chest CT scans of 1 mm slice thickness (C) and 3 mm slice thickness (F). CT, computed tomography.

External validation

External validation was performed on 334 paired scans (1 mm: 117; 3 mm: 117). The Bland-Altman plots showed

good agreement for both 1 and 3 mm scans [mean difference (95% limits of agreement) of 25 (−316.3 to 366.3) and 21.3 (−452.7 to 495.2)]. As for CAC category, the Kappa value was 0.80 (95% CI: 0.72–0.88) for the slice

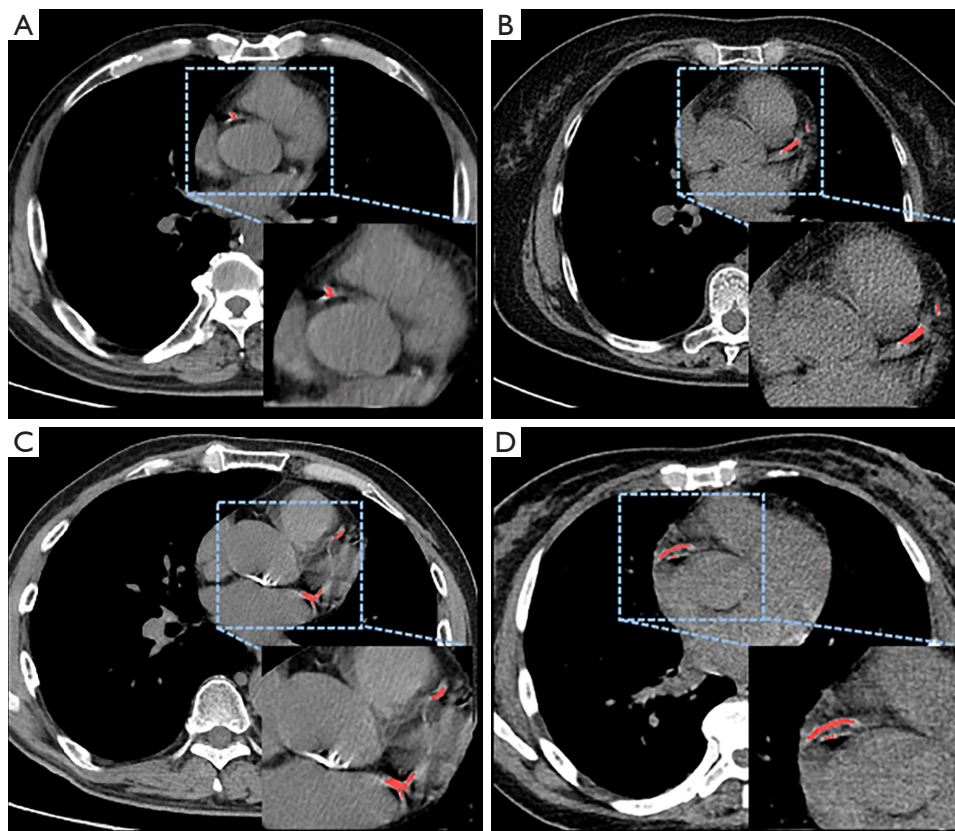


Figure 3 Example cases. Cross section images of non-gated chest CT. The image in the lower right corner shows an enlarged image of that inside the dotted blue box. Calcification plaques were correctly identified (A). Calcification plaques were correctly identified (B). Aortic calcification was distinguished from coronary calcification by the deep learning algorithm (C). Calcification in RCA was affected by cardiac motion (D). CT, computed tomography; RCA, right coronary artery.

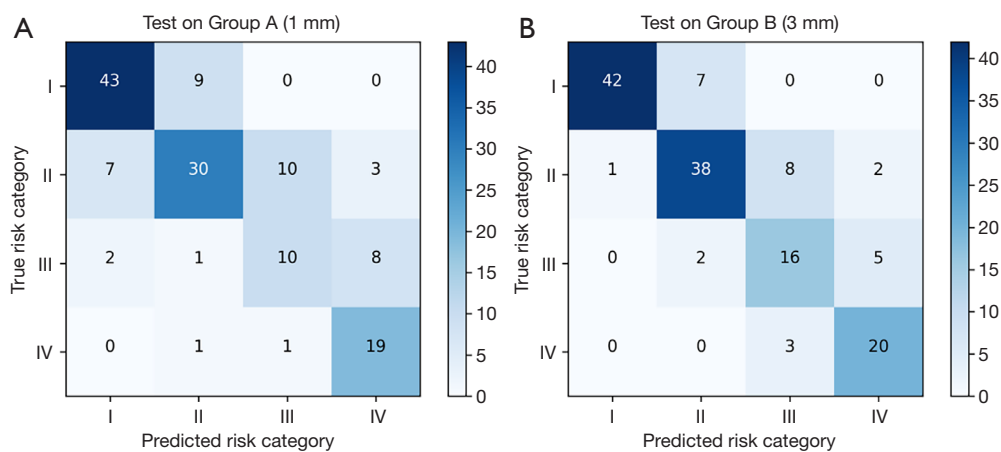


Figure 4 Classification confusion matrix. The classification confusion matrices of the coronary artery calcification categories. Truth risk categories based on the gold standard are depicted on the y-axis and prediction risk categories based on the deep learning algorithm are depicted on the x-axis of each matrix.

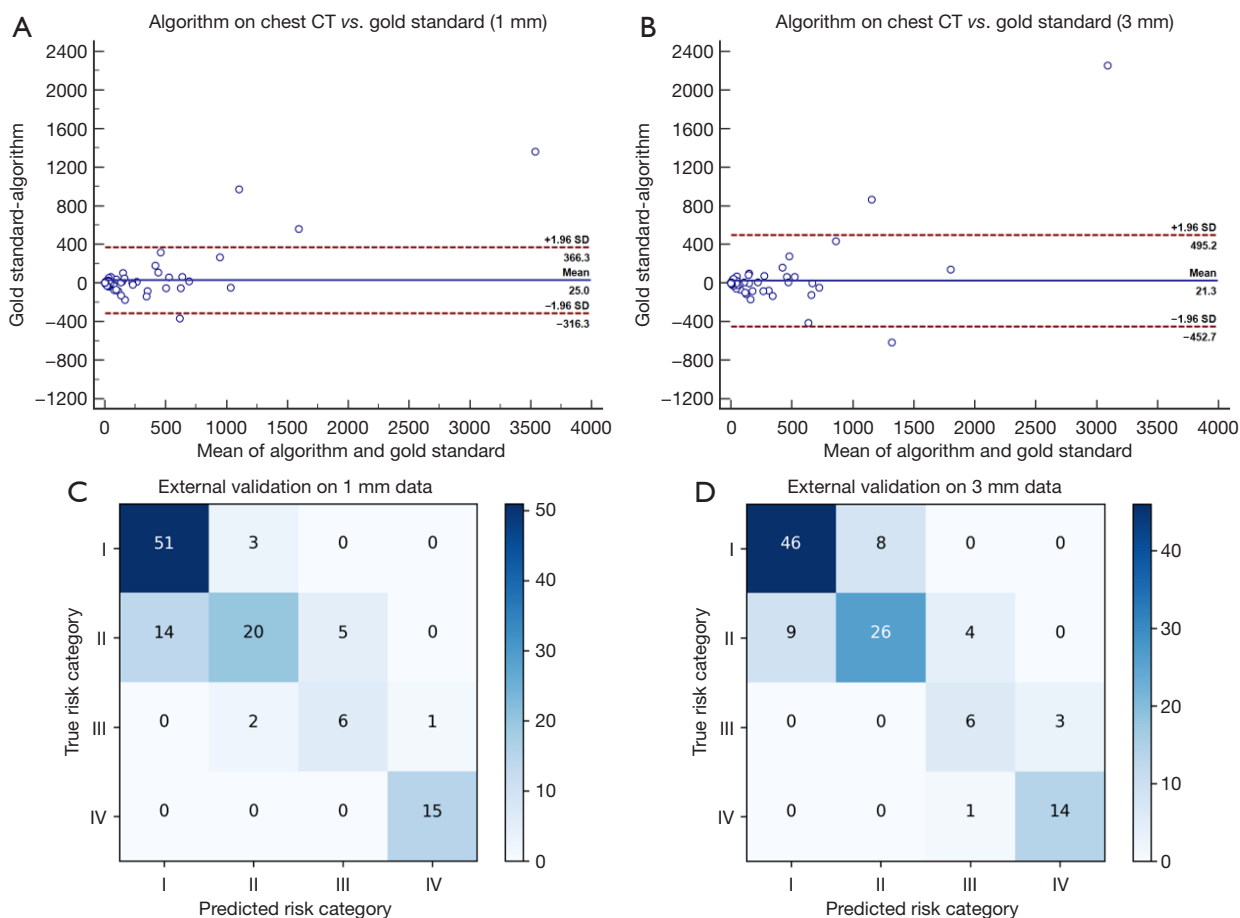


Figure 5 External validation results. The Bland-Altman plots of 1-mm-slice thickness (A) and 3-mm-slice thickness (B) chest CT. The classification confusion matrices of external validation on 1-mm-slice thickness (C) and 3-mm-slice thickness (D) chest CT. CT, computed tomography.

thickness of 1 mm and 0.80 (95% CI: 0.73–0.88) for the slice thickness of 3 mm. *Figure 5* shows the detailed Bland-Altman results and classification confusion matrices of the external validation.

Discussion

In the present study, we investigated a deep learning algorithm to automatically calculate the CAC score on non-gated, non-contrast chest CT of different slice thicknesses. The results showed that the algorithm was reliable and accurate for coronary calcification detection, quantification, and cardiac risk stratification on both 1-mm- and 3-mm-slice thickness chest CT. External validation also yielded good results. This method can save time spent on manual calculations and complete cardiac risk assessments while

simultaneously screening for chest diseases on chest CT.

The CAC score derived from cardiac CT remains the gold standard in clinical practice. Previous studies have demonstrated good consistency between manual calcification scores derived from chest CT and dedicated calcium scoring CT. Hutt *et al.* (24) investigated the reliability of non-gated CT for CAC screening in 185 individuals, however, there were 74 negative cases and 44 severe calcification cases, accounting for more than half of the data. Wu *et al.* (25) investigated 483 cases with 262 negative cases on low-dose, non-gated multi-detector CT, and the mean intertechnique scoring variability was 40–43%. The data in these studies contained too many negative cases (CAC score 0) or cases of severe calcification (CAC score ≥ 300), causing the consistency of the results to appear better than it actually was. The study by Budoff *et al.* (26)

involved 50 patients, representative of a small sample size. A meta-analysis showed that, although the agreement between the 2 techniques was relatively high ($\kappa = 0.89$), non-gated CT yielded an 8.8% false-negative calcification score and an underestimated high calcification score in 19.1% of cases (27). In our study, Bland-Altman analyses of the manual results on chest CT and the dedicated calcium scoring CT demonstrated the inherent differences between the 2 tests. From the perspective of automatic scoring using deep learning-based algorithms, few studies have adopted dedicated calcium scoring CT as a reference. van Assen *et al.* (16) examined 95 paired scans within 1.5 years and Eng *et al.* (17) examined 447 paired scans, with a maximum time interval of 2 years. CAC may show progression, and the accuracy of the deep learning algorithm may be affected. The time interval of 1 month in our study could avoid increases in the CAC score over time.

To the best of our knowledge, ours is the first study to compare the influence of different slice thicknesses on the deep learning-based automatic CAC scoring algorithm. Results showed that both 1 and 3 mm scans achieved good agreement with the gold standard, and chest CT with a slice thickness of 3 mm was slightly more accurate in CAC detection, with a sensitivity of 99% and a diagnostic accuracy of 94%. Previous studies have reported a sensitivity of 82% to 94% in detecting CAC (9-11,16,17). The sensitivity of our deep learning algorithm was higher than in previous studies, reducing the rate of missed diagnosis. The diagnostic accuracy was slightly higher than that of van Assen *et al.* (16), who reported an accuracy of 90%. The quantification of the CAC score on 3-mm-slice thickness CT was almost perfect in internal scans (ICC = 0.94) and good in external validation (ICC = 0.83). As for cardiac risk classification, our deep learning algorithm accurately classified 81% of internal and 79% of external cases, and the Kappa values were 0.82 and 0.80, respectively, which were higher than those of Cano-Espinosa *et al.* (11). The miscategorization rate was also in line with that of van Assen *et al.* (16) and Wolterink *et al.* (5). Most misclassified cases were within one category, with only 1% of cases showing more than one category in internal scans. Most misclassified cases were reclassified to higher categories, which hardly affected clinical decision making. This may be due to the increased volume of coronary calcium caused by artifacts. Only 1 (1%) false negative case and 6 (4%) underestimated cases were found on internal 3-mm-slice thickness chest CT, lower than previously reported (26).

Automatic CAC quantification on non-gated chest CT

is particularly cumbersome because of high noise, low resolution, and motion artifacts (13,28). The solutions to tackle this challenging task should be focused on the following: (I) accuracy, i.e., maintaining good consistency with the gold standard; (II) reliability, i.e., maintaining good consistency with manual results on chest CT; (III) easy interpretation, i.e., pixels that contribute to the calcium score are well classified in a single forward pass. Based on these considerations, we employed a two-stage segmentation method in this study. Our automatic scoring pipeline followed the identification and quantification style. Readers may refer to the work of Agatston *et al.* (18) on the accepted method of calculation of the calcification integral for quantification details, where the identification process is further discussed. As described previously (29), various tissues may be mistaken for calcified plaques in the surrounding environment of coronary arteries. Constructing a simple bounding box as an ROI is likely to be insufficient for subsequent CAC identification. However, our converged U-Net, which was trained with elaborate pixel-level annotated data, can predict a more precise ROI than previous methods when a new image is fed as the input. Given the heart ROI, many calcified plaques are still relatively small in the field of view, and segmenting such small objects is challenging due to the problem of heavy data imbalance. *Figure 6* shows the automatic coronary calcification identification process and *Figure 7* shows the flowchart of the post-processing phase using the deep learning algorithm.

Although our method obtained good overall performance, this study has some limitations. First, a chest CT-specific risk category was not investigated. Classification confusion matrices showed that our deep learning algorithm on chest CT tended to overestimate the risk category, and a chest CT-specific risk category may support the development of individualized treatment plans. Second, our method might fail for images with heavy motion artifacts or metal artifacts. Third, there were slightly more training data for the slice thickness of 3 mm than for 1 mm (170 *vs.* 150), which might have affected the result of the deep learning algorithm. Fourth, the sample size was too small to investigate the influence of CT scanner models or manufacturers on the results, and further studies are warranted on a larger scale to confirm our preliminary results.

In conclusion, our study proposed a reliable and accurate deep learning algorithm to automatically calculate coronary calcification scores and perform cardiovascular risk classification on non-gated, non-contrast chest CT.

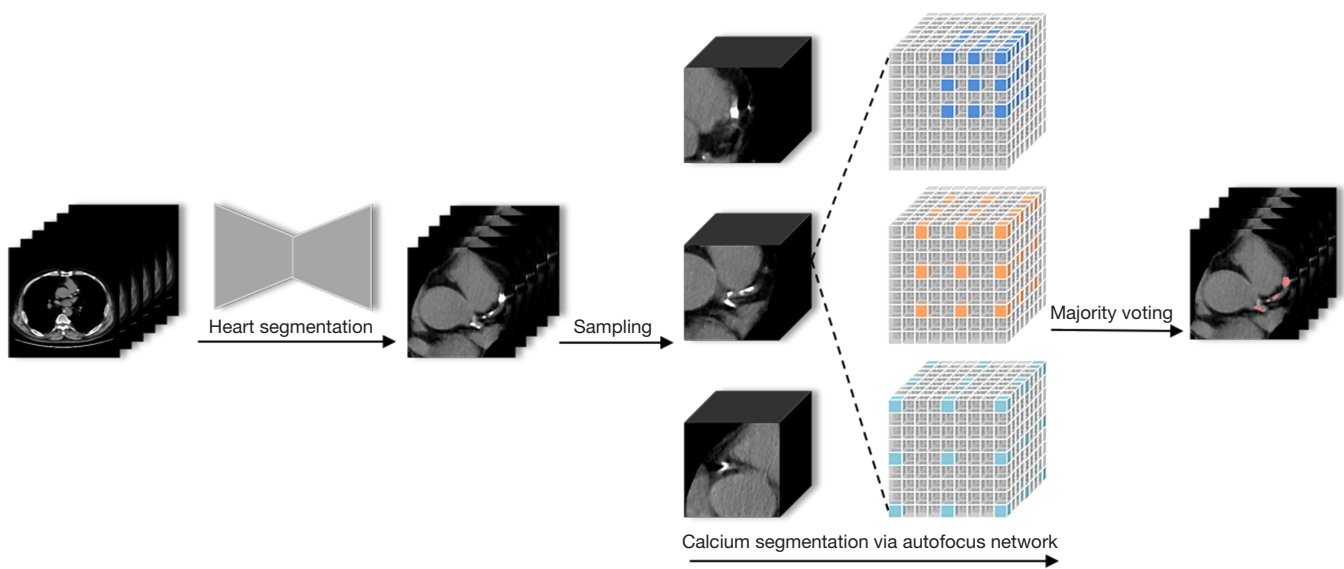


Figure 6 Automatic identification process. Automatic coronary calcification identification process from chest CT including the process of heart segmentation, sampling, calcium segmentation, and majority voting. CT, computed tomography.

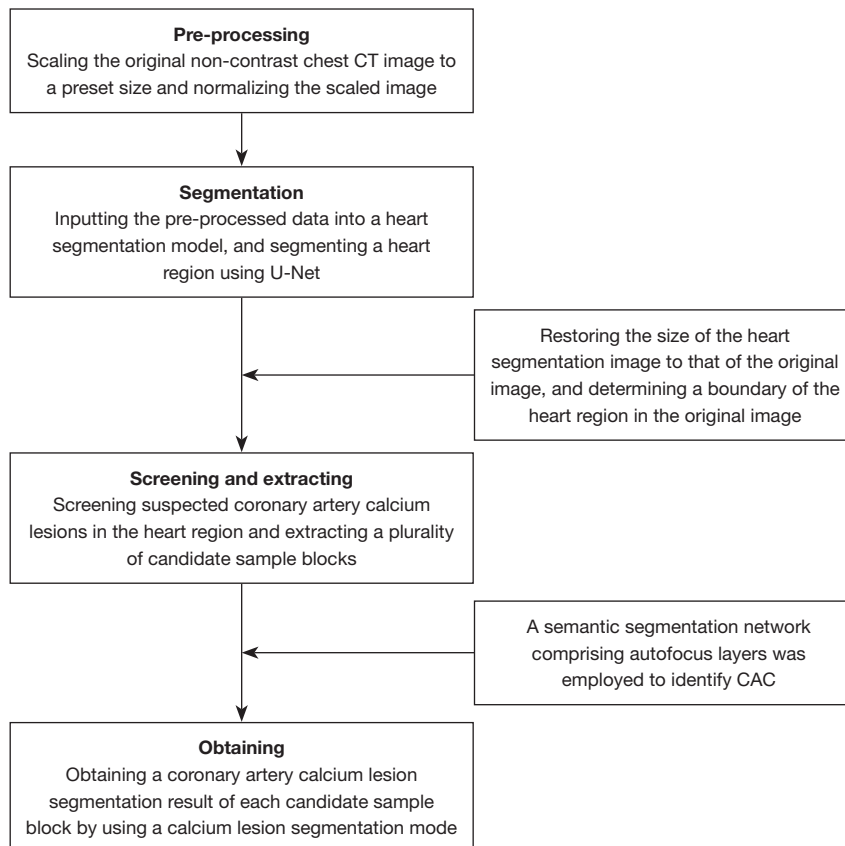


Figure 7 Flowchart of the post-processing phase using the deep learning algorithm. CT, computed tomography; CAC, coronary artery calcium.

The technique may be promising for clinical use, and may enable cardiovascular risk assessment to be undertaken while simultaneously screening for lung diseases on chest CT.

Acknowledgments

Funding: This work was supported by Beijing Natural Science Foundation (No. Z210013), CAMS Innovation Fund for Medical Sciences (CIFMS) (No. 2020-I2M-C-T-B-034), and the National Natural Science Foundation of China (No. 81873891).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-21-1017/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-21-1017/coif>). HG, MX, CX, ZJ, and YW have a patent (No. ZL 2020 1 1181703.0) issued. HG and MX are employed by Alibaba Group. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Institutional Ethics Committee of the Peking Union Medical College Hospital. Written informed consent was waived due to the retrospective nature of the study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, Greenland P. Coronary artery calcium score and risk classification for coronary heart disease prediction. *JAMA* 2010;303:1610-6.
- Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;139:e1082-143.
- Polonsky TS, Greenland P. Viewing the Value of Coronary Artery Calcium Testing From Different Perspectives. *JAMA Cardiol* 2018;3:908-10.
- Isgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging* 2012;31:2322-34.
- Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal* 2016;34:123-36.
- Lessmann N, van Ginneken B, Zreik M, de Jong PA, de Vos BD, Viergever MA, Išgum I. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions. *IEEE Trans Med Imaging* 2018;37:615-25.
- Takx RA, de Jong PA, Leiner T, Oudkerk M, de Koning HJ, Mol CP, Viergever MA, Išgum I. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLoS One* 2014;9:e91239.
- van Velzen SGM, Lessmann N, Velthuis BK, Bank IEM, van den Bongard DHJG, Leiner T, de Jong PA, Veldhuis WB, Correa A, Terry JG, Carr JJ, Viergever MA, Verkooijen HM, Išgum I. Deep Learning for Automatic Calcium Scoring in CT: Validation Using Multiple Cardiac CT and Chest CT Protocols. *Radiology* 2020;295:66-79.
- Wolterink JM, Leiner T, Takx RA, Viergever MA, Išgum I. Automatic Coronary Calcium Scoring in Non-Contrast-Enhanced ECG-Triggered Cardiac CT With Ambiguity Detection. *IEEE Trans Med Imaging* 2015;34:1867-78.
- Zeleznik R, Foldyna B, Eslami P, Weiss J, Alexander I, Taron J, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat Commun* 2021;12:715.
- Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ. Automated Agatston Score Computation in non-ECG Gated CT Scans Using Deep Learning. *Proc*

- SPIE Int Soc Opt Eng 2018;10574:105742K.
12. González G, Washko GR, Estépar RSJ. Deep learning for biomarker regression: application to osteoporosis and emphysema on chest CT scans. *Proc SPIE Int Soc Opt Eng* 2018;10574:105741H.
 13. de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Isgum I. Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. *IEEE Trans Med Imaging* 2019;38:2127-38.
 14. Cano-Espinosa C, González G, Washko GR, Cazorla M, José Estépar RS. On the Relevance of the Loss Function in the Agatston Score Regression from Non-ECG Gated CT Scans. *Image Anal Mov Organ Breast Thorac Images* (2018) 2018;11040:326-34.
 15. Xie X, Greuter MJ, Groen JM, de Bock GH, Oudkerk M, de Jong PA, Vliegenthart R. Can nontriggered thoracic CT be used for coronary artery calcium scoring? A phantom study. *Med Phys* 2013;40:081915.
 16. van Assen M, Martin SS, Varga-Szemes A, Rapaka S, Cimen S, Sharma P, Sahbae P, De Cecco CN, Vliegenthart R, Leonard TJ, Burt JR, Schoepf UJ. Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: A validation study. *Eur J Radiol* 2021;134:109428.
 17. Eng D, Chute C, Khandwala N, Rajpurkar P, Long J, Shleifer S, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. *NPJ Digit Med* 2021;4:88.
 18. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 1990;15:827-32.
 19. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, O'Leary DH, Tracy R, Watson K, Wong ND, Kronmal RA. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med* 2008;358:1336-45.
 20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015:234-41.
 21. Cai S, Tian Y, Lui H, Zeng H, Wu Y, Chen G. Dense-Unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quant Imaging Med Surg* 2020;10:1275-85.
 22. Qin Y, Kamnitsas K, Ancha S, Nanavati J, Cottrell G, Criminisi A, Nori A. Autofocus Layer for Semantic Segmentation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G. editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Cham: Springer, 2018:603-11.
 23. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. *International Conference for Learning Representations 2015*;arXiv:1412.6980v2.
 24. Hutt A, Duhamel A, Deken V, Faivre JB, Molinari F, Remy J, Remy-Jardin M. Coronary calcium screening with dual-source CT: reliability of ungated, high-pitch chest CT in comparison with dedicated calcium-scoring CT. *Eur Radiol* 2016;26:1521-8.
 25. Wu MT, Yang P, Huang YL, Chen JS, Chuo CC, Yeh C, Chang RS. Coronary arterial calcification on low-dose ungated MDCT for lung cancer screening: concordance study with dedicated cardiac CT. *AJR Am J Roentgenol* 2008;190:923-8.
 26. Budoff MJ, Nasir K, Kinney GL, Hokanson JE, Barr RG, Steiner R, Nath H, Lopez-Garcia C, Black-Shinn J, Casaburi R. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. *J Cardiovasc Comput Tomogr* 2011;5:113-8.
 27. Xie X, Zhao Y, de Bock GH, de Jong PA, Mali WP, Oudkerk M, Vliegenthart R. Validation and prognosis of coronary artery calcium scoring in nontriggered thoracic computed tomography: systematic review and meta-analysis. *Circ Cardiovasc Imaging* 2013;6:514-21.
 28. Olveres J, González G, Torres F, Moreno-Tagle JC, Carbajal-Degante E, Valencia-Rodríguez A, Méndez-Sánchez N, Escalante-Ramírez B. What is new in computer vision and artificial intelligence in medical image analysis applications. *Quant Imaging Med Surg* 2021;11:3830-53.
 29. Wu J, Ferns G, Giles J, Lewis E. A fully automated multi-modal computer aided diagnosis approach to coronary calcium scoring of MSCT images. *Proc. SPIE Medical Imaging: Computer-aided Diagnosis 2012*;8315:83152I.

Cite this article as: Xu C, Guo H, Xu M, Duan M, Wang M, Liu P, Luo X, Jin Z, Liu H, Wang Y. Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT. *Quant Imaging Med Surg* 2022;12(5):2684-2695. doi: 10.21037/qims-21-1017