Proxy-CAC: Learning Coronary Calcium Segmentation from Volume-Level Scores

Sandra Emi Sato Imuro[1], Salman Siddique Khan[1], Guha Balakrishnan[1], Mohammad Alkhaleefah[2], Sadeer Al-Kindi[2], and Ashutosh Sabharwal[1]

[1] Rice University, Houston TX 77005, USA
[2] Houston Methodist, Houston TX, 77030, USA
ses15@rice.edu

Abstract. Coronary Artery Calcification (CAC) is a key predictor of atherosclerotic cardiovascular risk and is typically quantified by the Agatston score, a non-differentiable metric derived by segmenting CAC from a patient's ECG-gated CT scan. Automated segmentations of coronary calcifications can enable streamlining of clinical workflows. However, real-world cardiac CT databases typically only store the volume-level Agatston score without also storing the underlying coronary calcium segmentations. To enable automated CAC segmentation and Agatston score computation, we propose Proxy-CAC, an end-to-end 3D U-Net framework for CAC quantification trained primarily with a loss function leveraging a differentiable Agatston score approximation, along with limited supervision over a small set of scans with per-voxel annotations. We trained and evaluated Proxy-CAC on an in-house, clinical dataset of 27,700 gated CT scans with associated Agatston scores from a large hospital. Proxy-CAC outperforms baseline models trained on segmentation masks at predicting Agatston scores by a substantial margin, increasing Cohen's Kappa values from 0.71 to 0.81 when stratifying patients into standard risk tiers. This work opens the door to automated, interpretable CAC risk prediction relying only on the limited, clinically-relevant supervision available in real-world clinical datasets.

Keywords: Coronary artery calcification · Differentiable Agatston scoring · Computed tomography · End-to-end

1 Introduction

Cardiovascular disease is the leading cause of death worldwide [11, 16]. Coronary Artery Calcification (CAC) has become an established, guideline recommended, imaging strategy for evaluating cardiovascular disease risk, providing direct evidence of the underlying plaque burden for a patient [16], which is crucial to guide personalized preventive therapies. The clinical standard for quantifying CAC is the Agatston score [2, 16], derived from low dose, non-contrast, cardiac CT scans.

The Agatston score is computed from a CT by a non-differentiable process: (1) identifying calcified lesions per 2D slice with tissue density >= 130 Hounsfield

Units (HUs), (2) assigning density-based weights to each lesion, (3) multiplying density weight by the lesion's area, and (4) summing lesion-wise contributions across slices for the full 3D volume [2]. We can define it as

$$S = \sum_{l=1}^{L} A_l \cdot w(H_{\max,l}), \tag{1}$$

where A_l is the lesion area and $w(\cdot)$ is a density weighting function based on the maximum attenuation (H_{max}) of each connected component within lesion l, and L denotes the number of total lesions in the 3D volume. The density weighting uses the standard Agatston breakpoints: $w = 1$ for $130 \leq H_{\max} < 200$, $w = 2$ for $200 \leq H_{\max} < 300$, $w = 3$ for $300 \leq H_{\max} < 400$, and $w = 4$ for $H_{\max} \geq 400$ HU [2].

Although quantifying Agatston score relies on identifying and segmenting lesions, in clinical practice only the final score is typically included in structured reports. Consequently, large real-world datasets of gated CT scans only come paired with Agatston scores [4]. While developing models that directly regress CTs to Agatston score is possible with these datasets [10, 18], the lack of intermediate dense segmentation outputs is not practical for clinicians who require interpretability and verifiability [13]. Hence, there is a large, untapped potential source of CT data for CAC evaluation due to their lack of dense annotation.

In this work, we formulate CAC quantification as a score-supervised semantic segmentation problem, learning voxel-level calcium predictions directly from a volume-level Agatston score. We propose Proxy-CAC, an end-to-end CT segmentation framework built on a base 3D U-Net [15] trained with a differentiable Agatston score proxy in its loss function. We replace non-differentiable operations of the Agatston score with a combination of smooth sigmoid-based relaxations [20]. Furthermore, to deal with potential stability issues of the training paradigm, we propose a two-phase training strategy. In the first phase, we assume a small dataset of expert-segmented CT scans with which we can initialize the weights of the U-Net. In the second stage, we train the model fully on a larger dataset of ECG-gated CT volumes with only Agatston score labels.

Existing deep learning approaches for CAC analysis follow two paths: per-voxel segmentation training, requiring expensive expert annotations [8, 12, 21, 17], and direct regression-based solutions that do not provide per-voxel inferences [18]. Other approaches based on semi-supervised and multi-task learning using unlabeled data [9, 13] have also been proposed. However, to our knowledge, no existing model learns to predict per-voxel CAC from final, real-world Agatston scores.

We developed and evaluated Proxy-CAC on a private dataset of 27,700 ECG-gated CT scans (with corresponding Agatston labels) from a large, urban hospital system. Our results demonstrate that incorporating a differentiable Agatston approximation allows for effective training from clinical score labels, improving CAC risk stratification, increasing Cohen's Kappa values from 0.71 to 0.81 while producing interpretable segmentation masks. This work demonstrates that

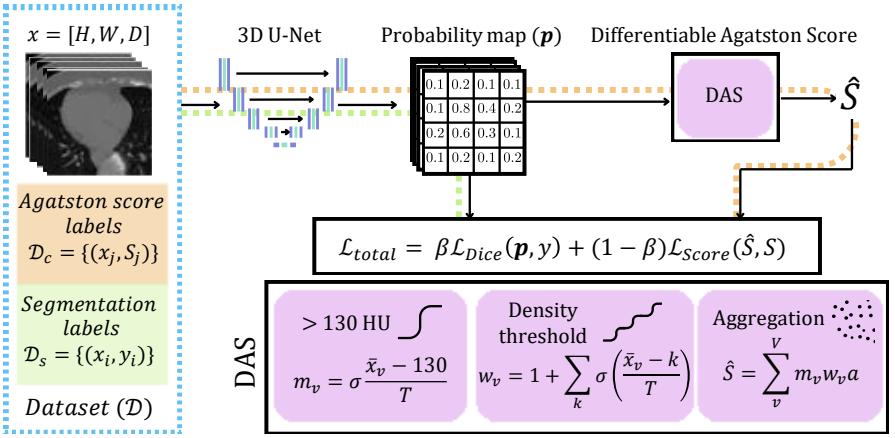

Fig. 1. Overview of the proposed Proxy-CAC pipeline for score-aligned coronary artery calcification (CAC) segmentation. The training is done in two phases: (1) a densely labeled dataset (D_s) with expert voxel-wise masks and (2) a large-scale clinically labeled dataset (D_c) containing scans paired only with clinical Agatston scores (S). A 3D U-Net generates a probability map (p), which is processed by a Differentiable Agatston Scoring (DAS) layer to produce a predicted score $\hat{S}$. The network is optimized via a hybrid loss function $\mathcal{L}_{total}$, where β is initially set to 1 to provide a strong anatomical anchor before incorporating global score-based gradients through backpropagation.

volume-level supervision, guided by clinical prior knowledge, can be used effectively to train clinically aligned, interpretable models at scale.

2 Methods

We assume an ECG-gated CT dataset $\mathcal{D} = \{\mathcal{D}_s, \mathcal{D}_c\}$ consisting of a small curated subset ($\mathcal{D}_s$) with voxel-level labels, and a large subset ($\mathcal{D}_c$) with volume-level Agatson score labels. $\mathcal{D}_s = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_s}$ consists of N_s CT volumes $\mathbf{x}$ and expert-annotated binary masks $\mathbf{y}$. $\mathcal{D}_c = \{(\mathbf{x}_j, S_j)\}_{j=1}^{N_c}$, is a high-volume repository of N_c gated CT scans where only the final Agatston score, S, is available. We assume $N_c \gg N_s$.

We aim to learn a functional mapping $f_\theta : \mathbf{x} \to \mathbf{p} \in [0,1]^V$, where V is the number of voxels in a scan, and $\mathbf{p}$ is a probability map quantifying per-voxel segmentation labels (see Fig.1). The segmentation backbone can be any model. In our experiments, we use a 3D U-Net [22]. Unlike standard segmentation models that end at the probability map, the core constraint we enforce in Proxy-CAC is to pass $\mathbf{p}$ through a custom Differentiable Agatston Score (DAS) module (see Section 2.1), creating an end-to-end pipeline where the final output is a scalar Agatston score, $\hat{S}$. Probability map $\mathbf{p}$ remains available as an intermediate

output to facilitate extra supervision from scans with dense annotations, and interpretability for clinical verification.

We optimize this network with two error signals using the stratified datasets described above. First, using $\mathcal{D}_s$, we minimize the discrepancy between $\mathbf{p}$ and the expert-annotated binary mask $\mathbf{y}$, providing a highly stable initial training phase for the network. Second, using $\mathcal{D}_c$, we minimize the discrepancy between a predicted Agatston score $\hat{S}$ and the clinical ground truth S. We now describe the DAS module and our training protocol.

2.1 Differentiable Agatston Score (DAS)

To optimize the network using the clinical scores in $\mathcal{D}_c$, the scoring function $DAS(\mathbf{x}, \mathbf{p})$ must be differentiable. In clinical practice, the Agatston Score is not differentiable (see Equation 1) [2].

We approximate the non-differentiable logic with a combination of sigmoid functions [20]. We calculate the predicted score $\hat{S}$ by the weighted sum across the volume:

$$\hat{S} = \sum_{v=1}^{V} m_v \cdot w_v \cdot a, \tag{2}$$

where v indexes each voxel, V is the total number of voxels in a volume, a is the cross-sectional area of each voxel, and m_v and w_v are as follows:

$$m_v = \sigma\left(\frac{\tilde{x}_v - 130}{T}\right), \tag{3}$$

$$w_v = 1 + \sum_{k \in B} \sigma\left(\frac{\tilde{x}_v - k}{T}\right). \tag{4}$$

Variable m_v approximates the 130 HU clinical threshold used to identify calcified voxels (see Fig. 1), where $\tilde{x}_v = x_v \cdot p_v$. represents the input voxel intensity in Hounsfield Units (HU) masked by the segmentation probability p_v. w_v approximates density weights $\{1, 2, 3, 4\}$ by summing responses at the clinical density breakpoints $B = \{200, 300, 400\}$ as explianed in the definition of Agatston score in Equation 1. Finally, T is a temperature parameter controlling the sharpness of the transition of the sigmoid function. We begin training with high T, creating a smooth gradient landscape, and anneal T toward a low number, approximating discrete thresholds that mimic the true Agatston scoring logic used in clinical practice.

2.2 Loss Function and Two-Phase Training

We stabilize the learning process through a two-phase approach, with the training loss function regulated by a trade-off parameter β:

$$\mathcal{L}_{total} = \beta \mathcal{L}_{Dice}(\mathbf{p}, \mathbf{y}) + (1 - \beta)\mathcal{L}_{Score}(\hat{S}, S), \tag{5}$$

where $\mathcal{L}_{Dice}$ is Dice score for per-voxel segmentation [7, 14] , and $\mathcal{L}_{Score}$ is a symmetric relative absolute error based on SMAPE to provide scale-invariance, which is essential for accurate risk stratification [3]:

$$\mathcal{L}_{Score}(S, \hat{S}) = \frac{\left|\hat{S} - S\right|}{|S| + |\hat{S}| + \epsilon}. \tag{6}$$

In Phase I, we train only on $\mathcal{D}_s$ and set $\beta = 1$, effectively using only $\mathcal{L}_{Dice}$. In Phase II, we set $\beta = 0$ and continue training the model on $\mathcal{D}_c$, while adjusting T to approximate discrete thresholds (see Equations 3 and 4).

3 Experiments and Results

We evaluated Proxy-CAC using data from two distinct datasets. For $\mathcal{D}_s$, we used the publicly available COCA-Coronary Calcium and Chest CT dataset [1] consisting of 787 ECG-gated CT scans with voxel-wise expert annotations for supervised training. For $\mathcal{D}_c$, we used a retrospective, real-world clinical repository of 27,700 gated CT scans with paired Agatston scores extracted from structured reports at Houston Methodist, a multi-site, multi-vendor hospital system. The use of institutional clinical data was approved or determined exempt by the Institutional Review Board (IRB Protocol No. PRO00037760). To ensure the fidelity of our approximation method, we validated the DAS outputs against the true Agatston scores within the COCA dataset, yielding a Pearson correlation of 0.996. For image preprocessing, we resampled all scans to $(0.3, 0.3, 3.0)$ mm resolution. Afterwards, we selected 40 slices from a heart-centered region of interest (ROI) using TotalSegmentator [19]. We downsampled all scans and masks to 256×256 resolution. We split $\mathcal{D}_s$ and $\mathcal{D}_c$ into training, validation, and test sets (70%/15%/15%) using ground truth CAC risk stratification to ensure similar distributions across splits.

Training: We implemented our model in PyTorch and trained on an NVIDIA A100-PCIE-40GB GPU. We used a 3D U-Net backbone with a feature map depth of $\{8, 16, 32, 64\}$. We first pretrained the model on $\mathcal{D}_s$. To address the class imbalance inherent in CAC segmentation, we filtered the training batches to ensure each batch contained at least 100 calcified voxels. We optimized the Dice loss ($\mathcal{L}_{dice}$) using the Adam optimizer with an initial learning rate of 10^{-2} and a step-decay schedule (step size 5, $\gamma = 0.5$). Second, we performed optimization on $\mathcal{D}_c$ using $\mathcal{L}_{score}$ with a learning rate of 10^{-4}. We decreased temperature T linearly from 10 to 0.5 during the first 50 epochs.

Evaluation metrics: We evaluated our model on a held-out test set using linear weighted Cohen's kappa [5, 6], (κ), based on the standard score stratification: 0 (no CAC), 1-10 (minimal), 11-100 (mild), 101-400 (moderate), >400 (severe) [2]. We report 95% boostrap CIs (B=1,000) throughout. To assess the agreement between predicted and ground-truth scalar values, we also evaluated the Pearson correlation coefficient R, and the coefficient of determination (R^2).

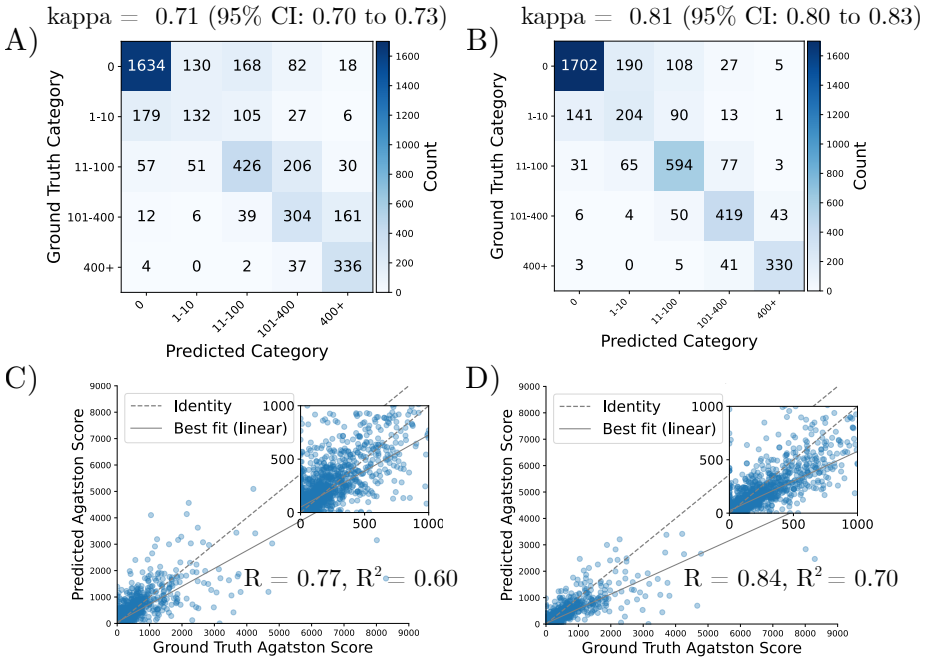

Fig. 2. Agatston score prediction and clinical risk stratification. (A) and (B) are confusion matrices for Agatston risk category classification. (A) displays the baseline performance of the 3D U-Net, while (B) shows the performance of the fine-tuned Proxy-CAC model. (C) and (D) are scatter plots illustrating the correlation between predicted and ground-truth Agatston scores. The fine-tuned Proxy-CAC model (D) demonstrates a higher Pearson correlation ($R = 0.84$, $R^2 = 0.70$) compared to the baseline 3D U-Net model (C) (blue, $R = 0.77$, $R^2 = 0.60$). Inset highlights clinically relevant CAC range (0-1000).

3.1 Agatston Score Estimation

We isolate the impact of using score-only supervision by evaluating the model at two stages: Phase I (equivalent to a standard supervised 3D U-Net trained on limited masks) and after the completion of Phase II (utilizing our differentiable agatston score (DAS) supervision). As illustrated in Figure 2, the transition to Phase II yielded a substantial improvement in clinical agreement, with the linear weighted Cohen's Kappa increasing from 0.71 (95% CI: 0.70-0.73) to 0.81 (95% CI: 0.80-0.83). The improvement in agreement was quantified as $\Delta\kappa = 0.10$ (95% CI: 0.09-0.11), with paired bootrap testing indicating statistical significance ($p<0.001$, B=1,000). The scalar correlation also demonstrated significant gains; Phase II reached a Pearson correlation of $R = 0.84$ ($R^2 = 0.70$), compared to $R = 0.77$ ($R^2 = 0.60$) at the end of Phase I.

Table 1. Comparison of CAC quantification performance. Our Proxy-CAC method (bold) leverages differentiable Agatston score supervision to outperform baseline and state-of-the-art methods in agreement metrics.

Method	**Supervision**	**Train → Test**	$\kappa \uparrow$	$R \uparrow$	$R^2 \uparrow$
AI-CAC	Mask	External → $\mathcal{D}_c$	0.73 [0.71, 0.74]	0.82	0.67
S2CAC	Mask + score[1]	External → $\mathcal{D}_c$	0.55 [0.53, 0.56]	0.83	0.68
3D UNet	Mask	$\mathcal{D}_s \rightarrow \mathcal{D}_c$	0.71 [0.70, 0.73]	0.77	0.60
Proxy-CAC (*)	**Score**	$\mathcal{D}_s + \mathcal{D}_c \rightarrow \mathcal{D}_c$	**0.81 [0.80, 0.83]**	**0.84**	**0.70**

Comparisons and benchmarking: To evaluate the real-world utility of our framework, we compared our model against two state-of-the-art architectures: AI-CAC [12] and S2CAC [13], using their publicly available code and pre-trained weights. All models were evaluated on the same hold-out test set from $\mathcal{D}_c$. Because AI-CAC was originally developed for full-chest non-gated CT scans, we performed inference for AI-CAC using the full-chest gated CT volumes from our test set. In contrast, S2CAC and our proposed model were evaluated using coned-down (cardiac-field-of-view) gated CT scans, consistent with their intended design. These results are summarized in Table 1. This comparison is meant to highlight a key practical constraint: Proxy-CAC shows improved performance compared to available models that cannot be fine-tuned on a dataset with only a scalar label, such as $\mathcal{D}_c$.

3.2 Segmentation Performance

The continuous probability maps (p) were binarized using a fixed threshold of 0.5 to produce final CAC segmentation masks. We provide a visual analysis of the segmentation masks in 3 comparing the output after Phase I and Phase II training. We observed two patterns: 1) Phase II outputs correctly remove extracardiac and non-coronary segmentations in our observed sample (Fig. 3 A-B). 2) Segmentation masks from Phase II often correctly identify the lesion, but may not cover the whole CAC (Fig.3 C-D).

While our large-scale dataset (D_c) lacks dense masks, a clinical expert annotated 50 random samples from the test set (10/category). We post-process predicted masks by identifying connected pixels within segmented regions above the clinical 130 HU threshold. We observed an increase in Dice score from 0.69 in Phase I to 0.79 in Phase II (0.55 to 0.61 without postprocessing). This jump demonstrates that scalar supervision provides superior lesion identification. Furthermore, Agatston scores derived from the refined masks achieve a Kappa of 0.83 [95% CI: 0.82, 0.84], demonstrating reliability of the produced masks for score computation.

[1] Scores were calculated from predicted segmented masks and were not the ground truth clinical report Agatston scores.

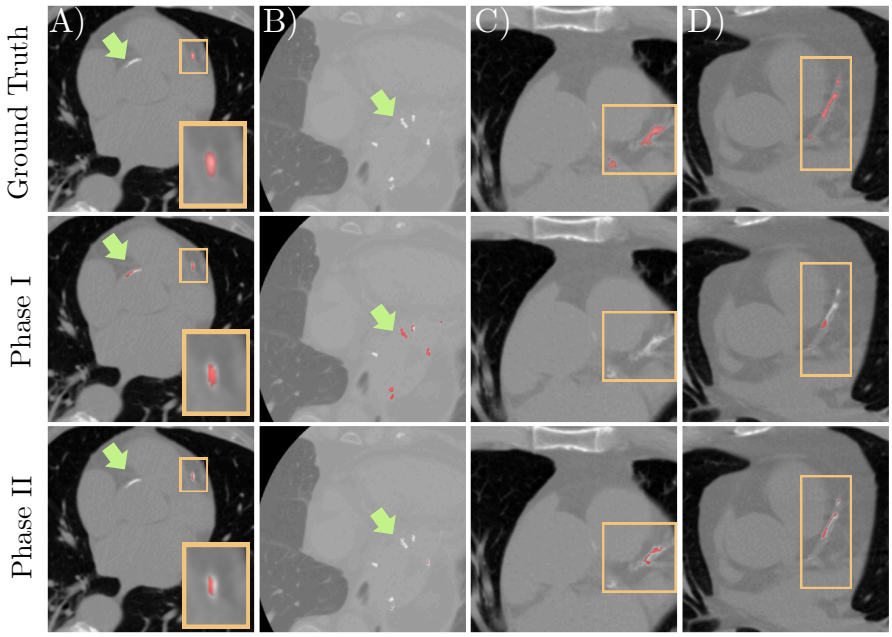

Fig. 3. Sample Proxy-CAC segmentation prediction results. We display a subset of four random patients (A-D) from the test set. Green arrows show non-coronary lesions correctly removed after Phase II (see A,B). Yellow boxes show changes in coronary calcium segmentations. C and D show an increase in true positive performance after Phase II.

4 Discussion

In this study, we introduced Proxy-CAC, a framework that uses Agatston score volume-level supervision to achieve high-performing coronary artery calcification (CAC) risk stratification. Central to this approach is a novel, differentiable approximation of the Agatston Scoring process. By using a combination of sigmoid functions to mimic the non-differentiable quantization steps of traditional scoring, we successfully integrated prior clinical knowledge into an end-to-end trainable network that produces interpretable segmentation masks. When evaluated on a private, in-house dataset of 27,700 CT scans, our model outperformed both standard 3D U-Net architectures and existing public repositories in predicting CAC risk categories.

Although we show promising results, our study has limitations, primarily the lack of ground-truth segmentation masks for large-scale quantitative evaluation. While qualitative reviews, such as those shown in Figure 3 indicate strong identification of calcified areas, further experimentation is required to refine these voxel-level outputs. In the future, our goal is to expand the scope of the model

and explore vessel-specific labeling and non-gated CT scans to broaden its clinical utility across a wider range of imaging protocols.

We emphasize that this challenge is not unique to CAC quantification but reflects a broader and ubiquitous limitation in clinical AI development. In routine practice, dense voxel-level annotations are rarely stored, as manual segmentation is labor-intensive, costly, and impractical at scale. Instead, large imaging repositories contain structured reports with final summary measures that are clinically meaningful but non-differentiable, such as Agatston scores. A hybrid strategy, such as the one proposed here, offers a pragmatic alternative: leverage a small, annotated dataset to anchor anatomical learning, while using large-scale repositories of reported outcomes. To solve this inherently unconstrained problem, our framework is engineered to impose structured constraints across two stages. Phase I provides a robust initialization for the segmentation model, preventing it from converging on non-anatomical degenerate solutions. Subsequently, Phase II uses the global score for refinement, with the DAS module incorporating clinical priors (such as the 130 HU threshold) to strictly limit the solution space to calcified voxels. This framework is modality agnostic and can be extended across imaging domains where structured clinical endpoints exist but dense annotations are unavailable, providing a generalizable pathway for trustworthy volume-level supervised learning in medicine.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Award No. 1648451. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Sadeer Al-Kindi is funded by NHLBI grant 5R01HL173186

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. COCA- Coronary Calcium and chest CT's | Center for Artificial Intelligence in Medicine & Imaging, https://aimi.stanford.edu/datasets/coca-coronary-calcium-chest-ct
2. Agatston, A.S., Janowitz, W.R., Hildner, F.J., Zusmer, N.R., Viamonte, M.J., Detrano, R.: Quantification of coronary artery calcium using ultrafast computed tomography. Journal of the American College of Cardiology **15**(4), 827–832 (Mar 1990). https://doi.org/10.1016/0735-1097(90)90282-t, place: United States
3. Armstrong, J.S.: Long-range forecasting : from crystal ball to computer. New York : Wiley (1985), http://archive.org/details/longrangeforecas0002arms
4. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A.V., Folsom, A.R., Greenland, P., Jacob, D.R., Kronmal, R., Liu, K., Nelson, J.C., O'Leary, D., Saad, M.F., Shea, S., Szklo, M., Tracy, R.P.: Multi-Ethnic Study of Atherosclerosis: objectives and design. American Journal of Epidemiology **156**(9), 871–881 (Nov 2002). https://doi.org/10.1093/aje/kwf113

5. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement **20**(1), 37–46 (Apr 1960). https://doi.org/10.1177/001316446002000104, https://doi.org/10.1177/001316446002000104
6. Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin **70**(4), 213–220 (1968). https://doi.org/10.1037/h0026256
7. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology **26**(3), 297–302 (1945). https://doi.org/10.2307/1932409, https://onlinelibrary.wiley.com/doi/abs/10.2307/1932409
8. Eng, D., Chute, C., Khandwala, N., Rajpurkar, P., Long, J., Shleifer, S., Khalaf, M.H., Sandhu, A.T., Rodriguez, F., Maron, D.J., Seyyedi, S., Marin, D., Golub, I., Budoff, M., Kitamura, F., Takahashi, M.S., Filice, R.W., Shah, R., Mongan, J., Kallianos, K., Langlotz, C.P., Lungren, M.P., Ng, A.Y., Patel, B.N.: Automated coronary calcium scoring using deep learning with multicenter external validation. npj Digital Medicine **4**(1), 88 (Jun 2021). https://doi.org/10.1038/s41746-021-00460-1, https://doi.org/10.1038/s41746-021-00460-1
9. Föllmer, B., Biavati, F., Wald, C., Stober, S., Ma, J., Dewey, M., Samek, W.: Active multitask learning with uncertainty-weighted loss for coronary calcium scoring. Medical Physics **49**(11), 7262–7277 (2022). https://doi.org/10.1002/mp.15870, https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.15870
10. González, G., Washko, G.R., Estépar, R.S.J., Cazorla, M., Cano Espinosa, C.: Automated Agatston score computation in non-ECG gated CT scans using deep learning. In: Angelini, E.D., Landman, B.A. (eds.) Medical Imaging 2018: Image Processing. p. 91. SPIE, Houston, United States (Mar 2018). https://doi.org/10.1117/12.2293681
11. Gulati, M., Levy, P.D., Mukherjee, D., Amsterdam, E., Bhatt, D.L., Birtcher, K.K., Blankstein, R., Boyd, J., Bullock-Palmer, R.P., Conejo, T., Diercks, D.B., Gentile, F., Greenwood, J.P., Hess, E.P., Hollenberg, S.M., Jaber, W.A., Jneid, H., Joglar, J.A., Morrow, D.A., O'Connor, R.E., Ross, M.A., Shaw, L.J.: 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation **144**(22), e368–e454 (Nov 2021). https://doi.org/10.1161/CIR.0000000000001029
12. Hagopian, R., Strebel, T., Bernatz, S., Myers, G.A., Offerman, E., Zuniga, E., Kim, C.Y., Ng, A.T., Iwaz, J.A., Nürnberg, L., Singh, S.P., Carey, E.P., Kim, M.J., Schaefer, R.S., Yu, J., Gentili, A., Aerts, H.J.: AI Opportunistic Coronary Calcium Screening at Veterans Affairs Hospitals. NEJM AI **2**(6), AIoa2400937 (May 2025). https://doi.org/10.1056/AIoa2400937, https://ai.nejm.org/doi/full/10.1056/AIoa2400937
13. Hao, J., Shah, N.S., Zhou, B.: S 2 CAC: Semi-supervised coronary artery calcium segmentation via scoring-driven consistency and negative sample boosting. Medical Image Analysis **107**, 103823 (Jan 2026). https://doi.org/10.1016/j.media.2025.103823, https://linkinghub.elsevier.com/retrieve/pii/S136184152500369X
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation (Jun 2016). https://doi.org/10.48550/arXiv.1606.04797

15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (May 2015). https://doi.org/10.48550/arXiv.1505.04597, arXiv:1505.04597 [cs]
16. Sandfort, V., Bluemke, D.: CT calcium scoring. History, current status and outlook. Diagnostic and Interventional Imaging **98**(1), 3–10 (Jan 2017). https://doi.org/10.1016/j.diii.2016.06.007
17. Sato Imuro, S.E., Siddique Khan, S., Ahmed, E.N., Sarda, D., Balakrishnan, G., Al-Kindi, S., Sabharwal, A.: Classifying Coronary Artery Calcification Risk using Radiomics from Non-Gated CT Scans. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–4 (Apr 2025). https://doi.org/10.1109/ISBI60581.2025.10980918, iSSN: 1945-8452
18. de Vos, B.D., Wolterink, J.M., Leiner, T., de Jong, P.A., Lessmann, N., Išgum, I.: Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. IEEE Transactions on Medical Imaging **38**(9), 2127–2138 (Sep 2019). https://doi.org/10.1109/TMI.2019.2899534
19. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology. Artificial intelligence **5**(5), e230024 (Sep 2023). https://doi.org/10.1148/ryai.230024, place: United States
20. Yang, J., Shen, X., Xing, J., Tian, X., Li, H., Deng, B., Huang, J., Hua, X.s.: Quantization Networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7300–7308. IEEE, Long Beach, CA, USA (Jun 2019). https://doi.org/10.1109/CVPR.2019.00748, https://ieeexplore.ieee.org/document/8953531/
21. Zeleznik, R., Foldyna, B., Eslami, P., Weiss, J., Alexander, I., Taron, J., Parmar, C., Alvi, R.M., Banerji, D., Uno, M., Kikuchi, Y., Karády, J., Zhang, L., Scholtz, J.E., Mayrhofer, T., Lyass, A., Mahoney, T.F., Massaro, J.M., Vasan, R.S., Douglas, P.S., Hoffmann, U., Lu, M.T., Aerts, H.J.W.L.: Deep convolutional neural networks to predict cardiovascular risk from computed tomography. Nature Communications **12** (2021), https://api.semanticscholar.org/CorpusID:231758750
22. Çiçek, , Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation (Jun 2016). https://doi.org/10.48550/arXiv.1606.06650