CS230

Prediction of Coronary Artery Disease Via Calcium Scoring of Chest CTs

Sahar Kazemzadeh, Manish Singh, Besher Ashouri, Samvel Gyurdzhyan,
Stanford University
saharkaz, msingh9, bashouri, sgyurdzh @stanford.edu

Abstract

Coronary artery calcium (CAC) scoring from chest computed tomography (CT) is used to assess risk of cardiovascular disease, which continues to be the leading avoidable cause of mortality in the United States. Currently, CAC scoring requires highly trained specialists to segment the calcification. This is expensive, rarely covered by insurance in the U.S., and poses a bottleneck in areas with scarcity of such trained specialists. To address these problems, we developed a deep learning system that segments coronary artery calcium from chest CTs. The segmentation achieved a Dice coefficient of 0.739 on the test set. From the predicted segmentations we produced CAC scores using the Agatston convention and found a high level of concordance with scores derived from the manual segmentations. Our results demonstrate the clinical value of a deep learning system for the automated prediction of cardiovascular events.

1 Introduction

Cardiovascular disease (CVD) is the leading avoidable cause of mortality in the United States[1]. While there exist effective pharmaceutical interventions and lifestyle modifications to treat CVD, identifying those who would benefit from such interventions continues to be a challenge[2]. Age and gender, as traditional risk markers, have a limited prognostic value for cardiovascular disease in humans. As a result, considerable advancements in the prediction and categorization of cardiovascular risk on an individual basis are necessary[3].

For several centuries, the association between vascular calcification and atherosclerotic disease has been recognized. The quantification of coronary artery calcium (CAC) was pioneered in 1990 by Arthur Agatston and it remains one of the strongest predictors for adverse cardiovascular events[4]. First, the cardiac arteries captured via a gated computer tomography (CT) scan are segmented, and then calcification in those arteries is quantified by an Agatston score, which is the summed score of all lesions in the coronary arteries, weighted for the density of the calcium.[5–8]. Gating refers to an acquisition technique that triggers a scan during a specific portion of the cardiac cycle in order to obtain high-quality scans void of arterial pulsations that can interfere with the read of the scan.

Currently cardiac segmentation requires highly specialized skills and is labor intensive, which presents a bottleneck for coronary artery calcium screening. We aim to develop a deep learning algorithm that can quantify an Agatston score from a gated CT to enable widespread screening in areas with shortages of individuals trained for segmentation.

2 Related work

In CT scans, calcification presents itself as dense white lesions; this presentation is not limited to calcification, and can also denote artery walls, bones, and foreign objects such as stents. Furthermore, CT scans show several structures that can contain calcification, including the lungs, bones, and both coronary and non-coronary arteries [TODO doctors: please verify]. To do this task well a model must be able to distinguish between calcification and other dense lesions, and localize calcification only in the four coronary arteries.

Prior work on automatic calcium scoring can be divided into three categories: rule-based approaches, conventional machine learning approaches and, most recently, deep learning approaches.

Early rule-based approaches identified CAC based on manually identified features such as location, shape, size, texture, and intensity, and then used a two-stage classification approach of k-nearest neighbors and support vector machines[9,10]. To determine location features, the CT images were registered to atlas images to help create probabilistic maps of CAC locations. The atlas images were made from tens of CTs with annotated coronary arteries[9,10]. This scheme is labor-intensive, time-consuming and has produced only average performance.

The second wave of automatic calcium scoring was done using improved machine-learning approaches. Instead of relying heavily on feature selection, randomized classifier trees were used to efficiently reduce dimensionality[11], and affine registration replaced deformable image registration, which cut computational complexity[12].

Deep-learning approaches represent the majority of recent work on automatic calcium scoring. Zelesnik et al[13] chained together 3 models for this task: heart localization, heart segmentation, and finally calcium segmentation. This achieved high performance, at the cost of being potentially difficult to productionize in real-time in low-resource settings due to the required computational overhead. More recently Eng et al were successful with a one shot approach.[14] We took a one shot approach as well.

3 Dataset and Features

We used the Stanford AIMI COCA dataset[15], which was collected by Stanford Hospital and Clinics. There were a total of 789 gated CT scans from 789 patients, with on average 50 512x512 images per scan. There were 36,411 total images, among which 3,656 contained at least one coronary artery calcification object. Each calcification object was also categorized as belonging to one of four coronary arteries. There were a total of 6,211 calcified objects, with an average pixel count of 119, such that they occupied on average only 0.04% of the image. This poses a challenging segmentation task; the challenge is amplified by a need to detect calcification in the 4 coronary arteries, and not elsewhere in the heart, lungs, or other visible organs.

3.1 Data Preprocessing

We implemented a data ingestion pipeline that parsed the calcification metadata from XMLs and stored them in a Pandas DataFrame[16] from which we derived binary masks. The image indexing of the metadata was unexpectedly in reverse order from the image indexes in the filenames, which sent us on a fun investigative adventure as we uncovered the issue. We eliminated 5 patients due to malformed calcification metadata. We randomly partitioned our patients into a train-validation-test split of 80%-10%-10%.

The dataset did not contain the ground truth calcification scores, which we derived from the segmentation masks as we describe later on.

4 Methods

The goal was to create binary segmentation masks that denoted regions of calcification in the coronary arteries, and to then compute Agatston scores from the masks that denote the level of coronary artery calcification. The more accurate the segmentation mask, the more accurate and thus clinically useful the Agatston score. Because arterial classification of the calcified object can be clinically useful, we also experimented with adding a classification head at the top of the network.

4.1 Baseline Model

For our baseline, we trained a model with U-Net architecture, which has been shown to work well for medical image segmentation[17], to which we added batch normalization. We also used dice loss, the Adam optimizer, and learning rate of 1e-4. We didn't have a sufficient number of volumes to take a volumetric approach, and thus treated this as a 2D image segmentation task.

4.2 Baseline to Improved

We faced two types of class imbalance problems: (1) we had a small fraction (10%) of images with at least one coronary artery calcification, and (2) of these images, the calcification occupied only 0.04% of the image.

4.2.1 Class Balance

To address the first type of class imbalance, we experimented with different positive image upsampling factors. Upsampling the positives to a very high positives to negative ratio produced a very high Dice score on positive images, at the expense of performance on the negatives. Ultimately we found a ratio of approximately 1:1 positive to negative ratio proved best on the validation set, which corresponded to a positives upsampling factor of 8.

4.2.2 Loss

We initially used dice loss, which measures the intersection over union between the predicted calcification class and ground truth calcification class, and is often used for image segmentation. We found original dice loss sometimes had convergence issues so we used a modified dice loss[18] that squares the ground truth and predicted masks before summing them. Its cost function is shown below.

$$\text{C(Y, } \hat{Y}) = \alpha(1 - \frac{2(\hat{Y}_s^T Y_s)+\gamma}{||Y_s||_2^2+||\hat{Y}_s||_2^2+\gamma}) - (1-\alpha)\frac{1}{N}\sum_s \sum_{c=1}^{4} I(Y_s == 1) \odot Y_c \odot log(\hat{Y}_c)$$

The γ smoothing factor avoids zero division. The first term is modified dice loss[18] and the second is categorical cross entropy loss for artery classification of segmented pixels. The indicator y_s in the second term is 1 when the pixel corresponds to a calcification. The full loss is a weighted sum of both the parts, where α is a hyper-parameter. In segmentation-only experiments, we excluded the second term. $\odot$ is element wise product.

To address the second type of class imbalance, we employed pixel-level focal loss, which is an improved version of cross-entropy loss that addresses class imbalance by assigning a higher loss to difficult examples and down-weights easy examples[19]. The cost function for binary focal loss is shown below, where c represents each class.

$$C(y, \hat{y}) = -\beta_t(1-p_c)^\gamma log(p_c) - (1-\beta_c)p_c^\gamma log(1-p_c)$$

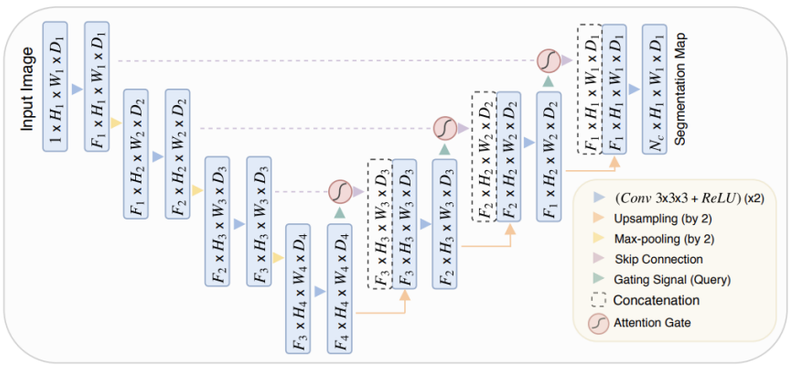

Figure 1: A block diagram of the Attention U-Net. Attention gates (AGs) filter the features propagated through the skip connections. Feature selectivity in the AGs is achieved by use of contextual information extracted in coarser scales. Reproduced from original publication[21]

Model Id	Model	Train DCE	Dev DCE	Test DCE
1	Baseline: Dice Loss U-Net	0.743	0.481	0.463
2	U-Net Focal Loss Upsample 4x	0.807	0.496	0.489
3	U-Net Dice Loss Trained on Positives	0.518	0.497	0.514
4	U-Net Dice Loss Upsample 4x	0.799	0.581	0.558
5	Custom Attention U-Net Focal Loss Upsample 8x	**0.914**	**0.813**	**0.739**

Table 1: Performance of the segmentation models across patient splits. The Custom Attention U-Net with focal loss and upsampling to produce a 1:1 ratio between positives and negatives outperformed our baseline model by 60%.

We employed a gamma of 2 and, to place a slightly higher penalty on false negatives than false positives, a β of 0.55.

We additionally experimented with focal tversky loss[20]., which is a combination of weighted dice loss and focal loss that is intended to address class imbalance in medical image segmentation, but focal loss proved empirically better for this task.

4.2.3 Architectures

Aside from standard U-Net, we tried a custom version of the Attention U-Net[21]. The Attention U-Net was introduced for medical image segmentation, and applies attention gates (AGs) that identify salient image regions and prune feature responses to preserve only the activations relevant to the target task. The architecture diagram is shown in Figure 1. It had 846,500 trainable parameters, while the standard U-Net had 34,523,969. We found that despite having about 2% of the trainable parameters as the standard U-Net, it outperformed the latter due to the efficacy of the attention gating. We noticed early signs of overfitting, so we introduced a novel component to this architecture by adding a dropout layer before the linear layer in the attention gate.

In all models we used batch normalization, and for all models except the baseline introduced a dropout rate of 20% to reduce overfitting to the train set. We eventually increased the learning rate to 1e-3, and used a batch size of 4 or 8 due to GPU memory constraints.

5 Results and Discussion

5.0.1 Model

The permutation that resulted in the best performing model was the Attention U-Net with focal loss and a positive upsample factor of 8. This achieved a Dice coefficient of 0.914 on the train set, 0.813 on the validation set, and 0.739 on the test set. The combination of the custom Attention U-Net and focal loss contributed most to this performance, which was a 60% improvement over our baseline.

5.0.2 False Positive and Negatives

Early checkpoints of the model had a lot of false positives and declared any dense white areas coronary artery calcification. In contrast the final model discerns between coronary artery and non-coronary artery calcification well. The aortic root calcification is appropriately ignored.

Our model captured some false positives and false negatives as demonstrated in Figure 2. In one case, it erroneously marked a coronary sinus wire from a resynchronization therapy device, a special type of catheter, as a protracted calcification. In another case, it captured a RCA stent as calcification. As for the false negatives they appear to have very small and faint calcification or are located near the aorta. Future work may prioritize purposefully up-sampling or augmenting images that include artefacts such as pacemakers and stents or images that have aortic calcification or aorta-adjacent calcification.

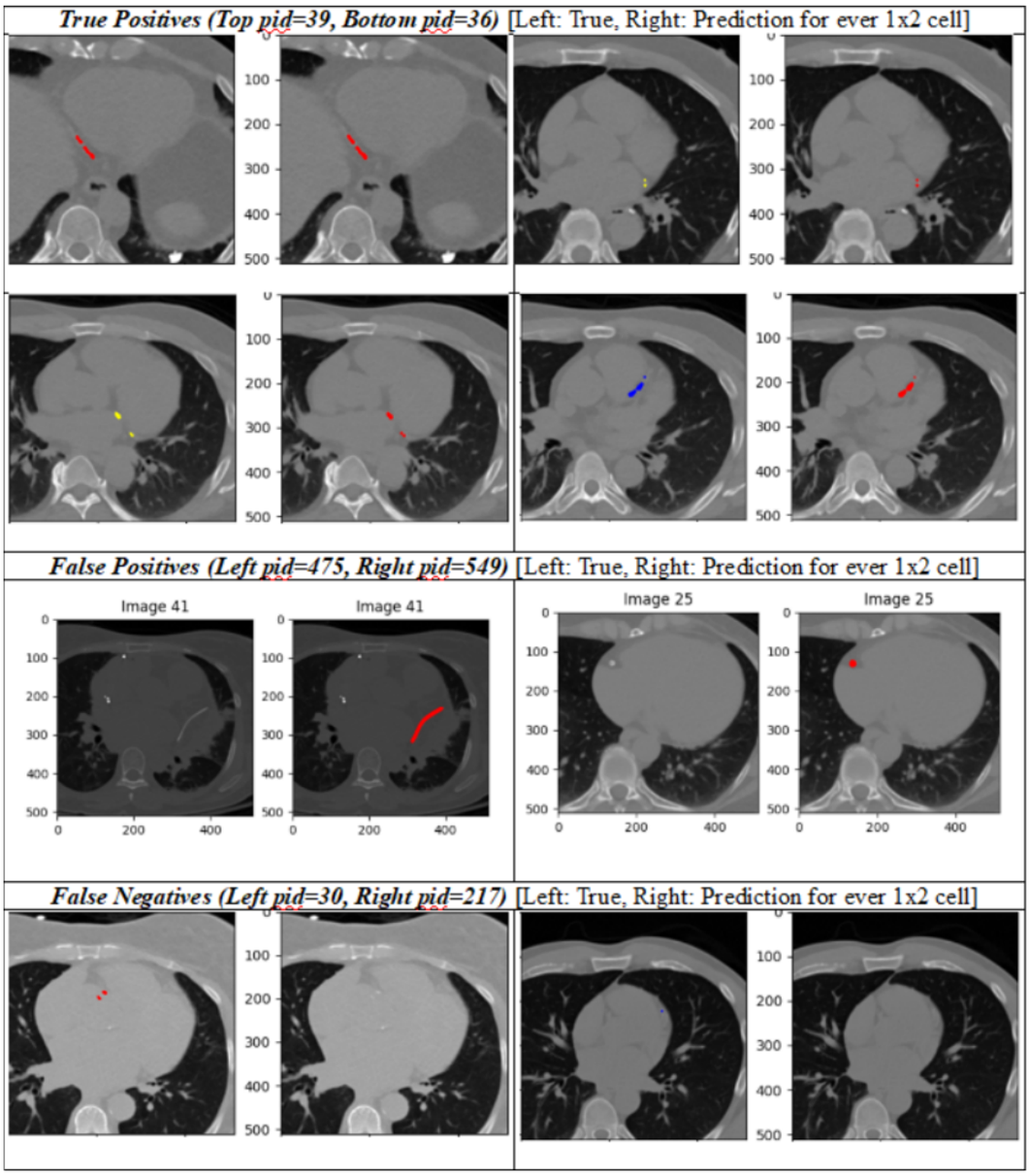

Figure 2: Example predictions from Model 3

α	Train DCE	Dev DCE	Dev Artery Class Accuracy
0.5	0.397	0.40	-
0.8	0.928	0.88	0.97
0.9	0.937	0.89	0.94
0.97	0.939	0.89	0.76
1.0	0.961	0.93	-

Table 2: Performance of the segmentation + artery classification models on positive images. These models were trained on only positives. Ultimately we focused exclusively on segmentation to maximize segmentation performance and thus clinical value.

5.0.3 Artery Classification

We trained some models to do both calcification segmentation and artery classification of the segmentations. The 4 artery classes were the left coronary artery, right coronary artery, left circumflex, and left anterior descending artery. For these experiments, we trained only on positive images, and achieved strong performance on both tasks when evaluating on positive images. We tuned the α hyperparameter in 4.2.2 to find the best loss weighting for the segmentation and classification components. Ultimately we shifted to focusing on segmentation exclusively to maximize segmentation performance and thus clinical value. Results are in Table 2.

5.0.4 Agatston Scores

From the ground truth and predicted segmentation masks, we computed Agatston scores to provide clinical comparison of our model with human raters[20]. The mean ground truth Agatston score for patients in our test set was 198 with a range of 2851, and the mean predicted Agatston score with Model 5 was 192. Furthermore,

based on the Agatston scores we bucketed patients into standard risk categories[4] and found a high level of concordance with ground truth categories, as shown in chart Figure 3. For example, our model correctly categorized 80% of the patients with a true risk category of 4 as falling into that bucket. The remaining 20% were categorized as level 5.

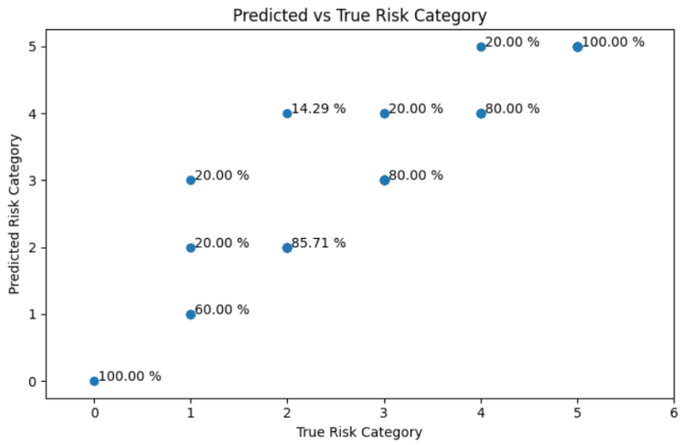

Figure 3: Computed risk category based on predictions from Model 3 vs ground truth. The Y axis gives the distribution of prediction for a particular X.

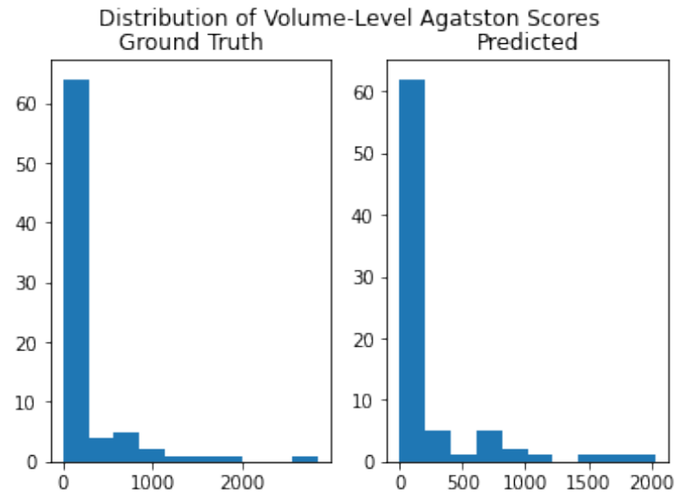

Figure 4: Distribution of volume-level Agatston scores between the ground truth and Model 5

6 Conclusion

Our results demonstrate the power of combining focal loss and Attention U-Net on the severely class imbalanced problem of coronary artery calcium segmentation. There is much potential in using deep learning to screen patients for heart disease in areas with a scarcity of technicians trained for CT segmentation of detection of coronary calcification.

6.0.1 Future Work

Deep learning algorithms can introduce tremendous value in the clinic by reducing unwanted variation in clinical practice, democratizing certain diagnostics to larger masses, and contributing to the detection of novel or traditionally undetectable disease states. While feasibility studies such as ours exist and are gaining traction in various domains of medicine, the adoption of deep learning algorithms in clinical practice is still marred with several barriers. Most of the work, as was ours, has been retrospective; it is imperative that prospective study designs be implemented to truly understand these algorithms, since performance is likely to be worse with real-world contemporaneous data that differs from that used in algorithm training. Additionally, randomized control trials must be implemented to deliver both a higher level of evidence and accordingly, facilitate the trust that will lead to adoption of AI in the bedside. Additional challenges exist with respect to generalization, and site-specific model training may be a requisite prior to adoption given systematic differences between health-systems and their respective populations.

Arguably more important in the medical space than any other is the problem of interpretability and explainability. Our model easily permits explainability because both the segmentation mask and Agatston score can be displayed to physicians. This type of design facilitates adoption in the clinic and fosters trust with both providers and patients.

7 Acknowledgement

The authors thank Ruta Joshi for guidance on this work, David Eng for patiently answering questions about the AIMI dataset and to Google Colaboratory for offering hassle-free, affordable access to the Tesla P100 when competitors made it oh-so-difficult to get access to a GPU.

8 Contributions

All members of the team regularly prepared for and attended a minimum of two meetings per week. Each member contributed valuable information to promote the advancement of the collective goals.

We implemented 2 end-to-end model training codebases.

Sahar Kazemzadeh implemented one of the codebases, did data preprocessing, trained models (including the search for the best architecture and loss functions), and did model evaluation including Agatston scoring.

Manish Singh implemented early codebase (https://github.com/msingh9/cs230-Coronary-Calcium-Scoring-) with scripts to visualize CT scan data and to be able to experiment and train models, trained unet model with modified dice loss as one of the final models, experimented with hyper parameters such as batch norm, drop out, intializers, and other loss functions to overcome convergence issues we faced early on. Manish Singh managed to train unet model with modified dice loss without any batchnorm and dropout layers, on his home machine with Titan GPU and did model evaluation to produce content for reports/videos.

Besher Ashouri and Samvel Gyurdzhyan contributed the background and literature research for the proposal, milestone and the final report. Contributed the calcium calculation code (Agatston) and the HU transformation code. Performed manual review and medical interpretation of every CT scan to capture mislabeled data (input and output). Contributed the slides for the final video.

References

[1] Dariush Mozaffarian, Emelia J. Benjamin, Alan S. Go, Donna K. Arnett, Michael J. Blaha, Mary Cushman, Sandeep R. Das, Sarah de Ferranti, Jean-Pierre Després, Heather J. Fullerton, Virginia J. Howard, Mark D. Huffman, Carmen R. Isasi, Monik C. Jiménez, Suzanne E. Judd, Brett M. Kissela, Judith H. Lichtman, Lynda D. Lisabeth, Simin Liu, Rachel H. Mackey, David J. Magid, Darren K. McGuire, Emile R. Mohler, Claudia S. Moy, Paul Muntner, Michael E. Mussolino, Khurram Nasir, Robert W. Neumar, Graham Nichol, Latha Palaniappan, Dilip K. Pandey, Mathew J. Reeves, Carlos J. Rodriguez, Wayne Rosamond, Paul D. Sorlie, Joel Stein, Amytis Towfighi, Tanya N. Turan, Salim S. Virani, Daniel Woo, Robert W. Yeh, and Melanie B. Turner. Heart Disease and Stroke Statistics—2016 Update. *Circulation*, 133(4):e38–e360, January 2016. doi: 10/bvft. URL `https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000350`. Publisher: American Heart Association.

[2] Yashashwi Pokharel, Fengming Tang, Philip G. Jones, Vijay Nambi, Vera A. Bittner, Ravi S. Hira, Khurram Nasir, Paul S. Chan, Thomas M. Maddox, William J. Oetgen, Paul A. Heidenreich, William B. Borden, John A. Spertus, Laura A. Petersen, Christie M. Ballantyne, and Salim S. Virani. Adoption of the 2013 American College of Cardiology/American Heart Association Cholesterol Management Guideline in Cardiology Practices Nationwide. *JAMA cardiology*, 2(4):361–369, April 2017. ISSN 2380-6591. doi: 10/gnbt2b.

[3] Ryan Poplin, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng, and Dale R. Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158–164, March 2018. ISSN 2157-846X. doi: 10/gdq7wp.

[4] A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano. Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American College of Cardiology*, 15(4):827–832, March 1990. ISSN 0735-1097. doi: 10/d239dr.

[5] Axel Schmermund, Dietrich Baumgart, Stefan Möhlenkamp, Paul Kriener, Heiko Pump, Dietrich Grönemeyer, Rainer Seibel, and Raimund Erbel. Natural History and Topographic Pattern of Progression of Coronary Calcification in Symptomatic Patients. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 21(3):421–426, March 2001. doi: 10/ct868z. URL `https://www.ahajournals.org/doi/10.1161/01.ATV.21.3.421`. Publisher: American Heart Association.

[6] Scott M. Grundy, Neil J. Stone, Alison L. Bailey, Craig Beam, Kim K. Birtcher, Roger S. Blumenthal, Lynne T. Braun, Ferranti Sarah de, Tommasino Joseph Faiella, Daniel E. Forman, Ronald Goldberg, Paul A. Heidenreich, Mark A. Hlatky, Daniel W. Jones, Jones Donald Lloyd, Pajares Nuria Lopez, Chiadi E. Ndumele, Carl E. Orringer, Carmen A. Peralta, Joseph J. Saseen, Sidney C. Smith, Laurence Sperling, Salim S. Virani, and Joseph Yeboah. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary. *Journal of the American College of Cardiology*, 73(24):3168–3209, June 2019. doi: 10/gfh4hc. URL `https://www.jacc.org/doi/abs/10.1016/j.jacc.2018.11.002`. Publisher: American College of Cardiology Foundation.

[7] Harvey S. Hecht, Paul Cronin, Michael J. Blaha, Matthew J. Budoff, Ella A. Kazerooni, Jagat Narula, David Yankelevitz, and Suhny Abbara. 2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: A report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology. *Journal of Thoracic Imaging*, 32(5):W54–W66, September 2017. ISSN 1536-0237. doi: 10/gnb3nh.

[8] Matthew J. Budoff, Rebekah Young, Gregory Burke, J. Jeffrey Carr, Robert C. Detrano, Aaron R. Folsom, Richard Kronmal, Joao A. C. Lima, Kiang J. Liu, Robyn L. McClelland, Erin Michos, Wendy S. Post, Steven Shea, Karol E. Watson, and Nathan D. Wong. Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: the multi-ethnic study of atherosclerosis (MESA). *European Heart Journal*, 39(25):2401–2408, July 2018. ISSN 1522-9645. doi: 10/cszt.

[9] Ivana Isgum, Mathias Prokop, Meindert Niemeijer, Max A. Viergever, and Bram van Ginneken. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Transactions on Medical Imaging*, 31(12):2322–2334, 2012. doi: 10.1109/TMI.2012.2216889.

[10] Rahil Shahzad, Theo van Walsum, Michiel Schaap, Alexia Rossi, Stefan Klein, Annick C. Weustink, Pim J. de Feyter, Lucas J. van Vliet, and Wiro J. Niessen. Vessel specific coronary artery calcium scoring. *Academic Radiology*, 20(1):1–9, January 2013. doi: 10.1016/j.acra.2012.07.018. URL https://doi.org/10.1016/j.acra.2012.07.018.

[11] Jelmer M. Wolterink, Tim Leiner, Richard A. P. Takx, Max A. Viergever, and Ivana Isgum. Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. *IEEE Transactions on Medical Imaging*, 34(9):1867–1878, September 2015. doi: 10.1109/tmi.2015.2412651. URL https://doi.org/10.1109/tmi.2015.2412651.

[12] Felix Durlak, Michael Wels, Chris Schwemmer, Michael Sühling, Stefan Steidl, and Andreas Maier. Growing a random forest with fuzzy spatial features for fully automatic artery-specific coronary calcium scoring. In *Machine Learning in Medical Imaging*, pages 27–35. Springer International Publishing, 2017. doi: 10.1007/978-3-319-67389-9_4. URL https://doi.org/10.1007/978-3-319-67389-9_4.

[13] Roman Zeleznik, Borek Foldyna, Parastou Eslami, Jakob Weiss, Ivanov Alexander, Jana Taron, Chintan Parmar, Raza M. Alvi, Dahlia Banerji, Mio Uno, Yasuka Kikuchi, Julia Karady, Lili Zhang, Jan-Erik Scholtz, Thomas Mayrhofer, Asya Lyass, Taylor F. Mahoney, Joseph M. Massaro, Ramachandran S. Vasan, Pamela S. Douglas, Udo Hoffmann, Michael T. Lu, and Hugo J. W. L. Aerts. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nature Communications*, 12(1):715, January 2021. ISSN 2041-1723. doi: 10/gk2twp. URL https://www.nature.com/articles/s41467-021-20966-2. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computer science;Interventional cardiology;Machine learning;Outcomes research;Risk factors Subject_term_id: computer-science;interventional-cardiology;machine-learning;outcomes-research;risk-factors.

[14] Chute C. Khandwala N. et al Eng, D. Automated coronary calcium scoring using deep learning with multicenter external validation. 2021. URL https://doi.org/10.1038/s41746-021-00460-1.

[15] Stanford University Stanford AIMI. Coca - coronary calcium and chest cts, 2021. data retrieved from Stanford AIMI, https://stanfordaimi.azurewebsites.net/datasets/e8ca74dc-8dd4-4340-815a-60b41f6cb2aa.

[16] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://doi.org/10.1007/978-3-319-24574-4_28.

[18] Yuxian Meng Junjun Liang Fei Wu Xiaoya Li, Xiaofei Sun and Jiwei Li. Dice loss for data-imbalanced nlp tasks. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 465–476*, 2020.

[19] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, February 2020. doi: 10.1109/tpami.2018.2858826. URL https://doi.org/10.1109/tpami.2018.2858826.

[20] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, April 2019. doi: 10.1109/isbi.2019.8759329. URL https://doi.org/10.1109/isbi.2019.8759329.

[21] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.