SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

# Deep Learning for Automated Aortic Valve Calcium Scoring on Non-Gated Chest CT

Chanon Chantaduly, BA, University of California, Irvine; Shawn Sun, MD; Justin Grant, MD; Long-Co Nguyen, MD; Parker Rushworth, MD; Raffi Hagopian, MD; Jennifer Xu, MD; Peter D. Chang, MD

## Introduction/Background

Timely diagnosis of aortic valve (AV) stenosis is critical, yet current methods like echocardiography or gated CT are too resource-intensive for widespread screening. To address this gap, we developed and validated a deep learning model for fully automated opportunistic detection of AV stenosis on routine non-gated chest CTs.

## Methods/Intervention

This retrospective study utilized a cross-validation cohort of 706 consecutive non-gated CTs and a hold-out test cohort of 60 non-gated CTs with echocardiograms. AV leaflet calcification was manually annotated with segmentation masks. A fully automated preprocessing pipeline leveraging the DeepATLAS foundation model was designed to iteratively generate an input 3D volume cropped to the aortic root (Figure 1). The CNN model for AV calcification detection was implemented as a 3D encoder-decoder (U-Net) with 32 layers and 16M+ parameters, optimized using both focal and soft Dice loss. Model predictions and human annotations were converted to Agatston scores and mapped to stenosis grades using established gender-normalized thresholds.

## Results/Outcome

AI-generated calcium scores demonstrated high concordance with human ground-truth, achieving a Pearson correlation of 0.926, a Dice score of 0.713, and a mean absolute Agatston score error of 20.1 (IQR, 0.0-98.3). For stenosis classification (Table 1), the model achieved a weighted F1-score of 0.940 and an intraclass correlation coefficient (ICC) of 0.906 [CI, 0.89-0.92] for severe vs. non-severe cases. Out of 706 patients, the model identified 28/30 of those with severe stenosis, with the remaining two classified as moderate. In the test set with echocardiography ground-truth, the AI was superior or equivalent to human annotators for identifying severe stenosis (AUC: 0.901-0.949 [AI] vs. 0.873-949 [human]) and for differentiating severe from moderate disease (AUC: 0.788-0.821 [AI] vs. 0.727-0.821 [human]).

## Conclusion

Our model provides automated predictions of AV stenosis from non-gated CTs with high accuracy compared to expert and echo ground-truth. Even in the setting of population screening with low baseline prevalence, less than a quarter of predictions are false positives despite over 93% sensitivity in identifying severe disease. This solution represents a promising method for opportunistic early detection of aortic stenosis.

## Statement of Impact

AI-enabled opportunistic detection of severe aortic stenosis accelerates patient access to life-saving interventions.

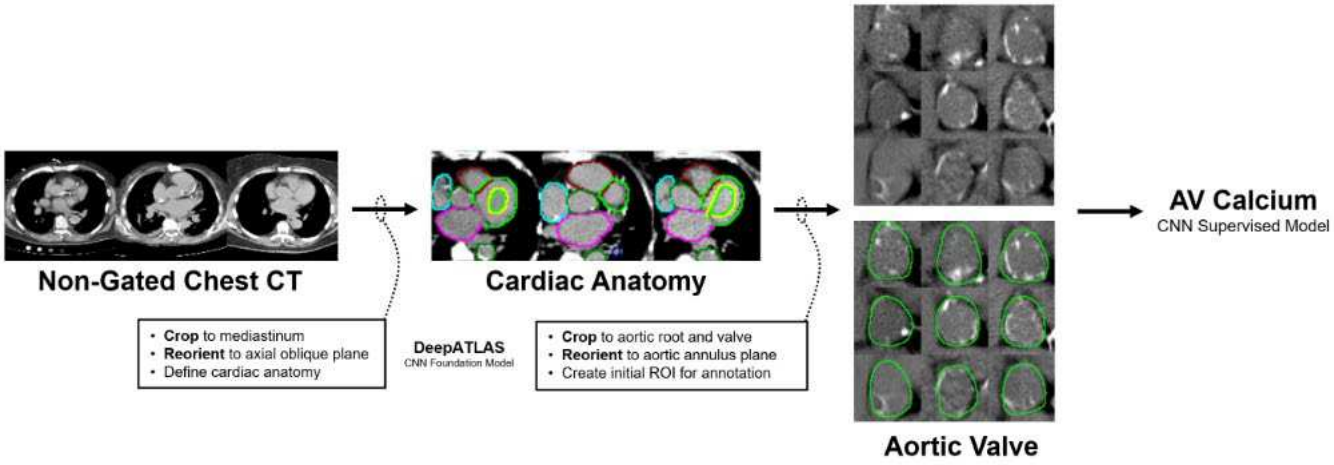


Figure 1. Overview of Deep Learning Modeling Strategy. The DeepATLAS CNN-based foundation model is used to localize relevant cardiac anatomy and sequentially process raw CT data. First, key cardiac structures (multi-color outlines) are identified from routine non-gated chest CT exams after cropping to the mediastinum. Second, the aortic root and valve are delineated (green outline). Third, key landmarks are used to define an oblique plane parallel to the aortic annulus. The final data volume is resampled to a uniform matrix shape and used as input to CNN model training.

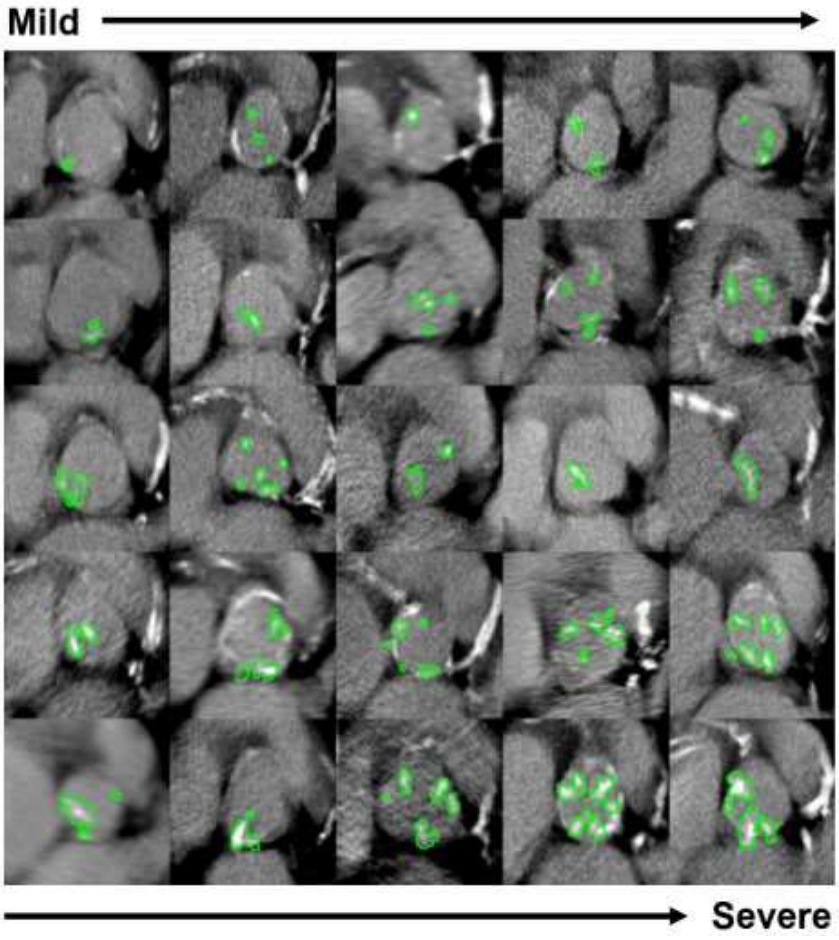


Figure 2. AI Generated Predictions for Aortic Valve Calcification. Deep learning model predictions (green) are shown overlaid upon 3-mm mean intensity projections in the region of the aortic root. Notice that the model can successfully identify the more central regions of leaflet calcification and ignore other distracting areas of high density for example along the aortic annulus and surrounding coronary arteries.

| TARGET / OUTPUT | Mild | Moderate | Severe | SUM |
|---|---|---|---|---|
| Mild | 604<br>85.55% | 24<br>3.40% | 5<br>0.71% | 633<br>95.42%<br>4.58% |
| Moderate | 10<br>1.42% | 29<br>4.11% | 4<br>0.57% | 43<br>67.44%<br>32.56% |
| Severe | 0<br>0.00% | 2<br>0.28% | 28<br>3.97% | 30<br>93.33%<br>6.67% |
| SUM | 614<br>98.37%<br>1.63% | 55<br>52.73%<br>47.27% | 37<br>75.68%<br>24.32% | 661 / 706<br>93.63%<br>6.37% |

| | |
|---|---|
| Accuracy | 0.9363 |
| Misclassification Rate | 0.0637 |
| Macro-F1 | 0.7988 |
| Weighted-F1 | 0.9401 |

| Class Name | Precision | 1-Precision | Recall | False Negative Rate | F1 score | Specificity (TNR) | False Positive Rate (FPR) |
|---|---|---|---|---|---|---|---|
| Mild | 0.9837 | 0.0163 | 0.9542 | 0.0458 | 0.9687 | 0.8630 | 0.1370 |
| Moderate | 0.5273 | 0.4727 | 0.6744 | 0.3256 | 0.5918 | 0.9608 | 0.0392 |
| Severe | 0.7568 | 0.2432 | 0.9333 | 0.0667 | 0.8358 | 0.9867 | 0.0133 |

Table 1. Summary of Model Performance Statistics. For all analysis, human and AI algorithm predictions are converted to standardized Agatston scores and mapped to stenosis grade based on published thresholds for males (mild 2000) and females (mild < 400, moderate 400-1300, severe > 1300).


## Keywords

Aortic Stenosis; Opportunistic Screening; CT; Convolutional Neural Network; Supervised Deep Learning; Cardiology