SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

Concordance of Qualitative Aortic Descriptors with Diameter and Aortic Height Index: An LLM-Based Analysis of 1,000 CT Angiography Reports

Trent J. Levy, MD; Perelman School of Medicine, University of Pennsylvania;
Tessa S. Cook, MD, PhD, CIIP, FSIIM

Introduction/Background
Radiology reports convey aortic risk through unstandardized descriptors, while guidelines increasingly endorse height-indexed metrics; the ACC/AHA cites an Aortic Height Index (AHI) of ≥3.21 cm/m as an alternate surgical threshold. Whether report terminology reflects height-indexed risk, and whether AHI improves on raw diameter, has not been tested at scale.

Methods/Intervention
Retrospective analysis of 1,000 consecutive CT angiography reports at a single academic center. An LLM pipeline extracted maximal native aortic diameter, segment, and qualitative descriptors, with negation-scope handling and exclusion of history-section terms, prior-study values, and stent/graft measurements. Manual audit of extracted measurements identified 11 erroneous extractions (1.7%), which were excluded. Descriptors were collapsed into three ordered severity tiers; morphology terms were analyzed separately. Association with diameter and AHI was assessed by Kruskal–Wallis, Spearman correlation, and ROC analysis with bootstrapped AUC comparison. Analytic cohort: 278 reports with diameter, height, and a severity descriptor.

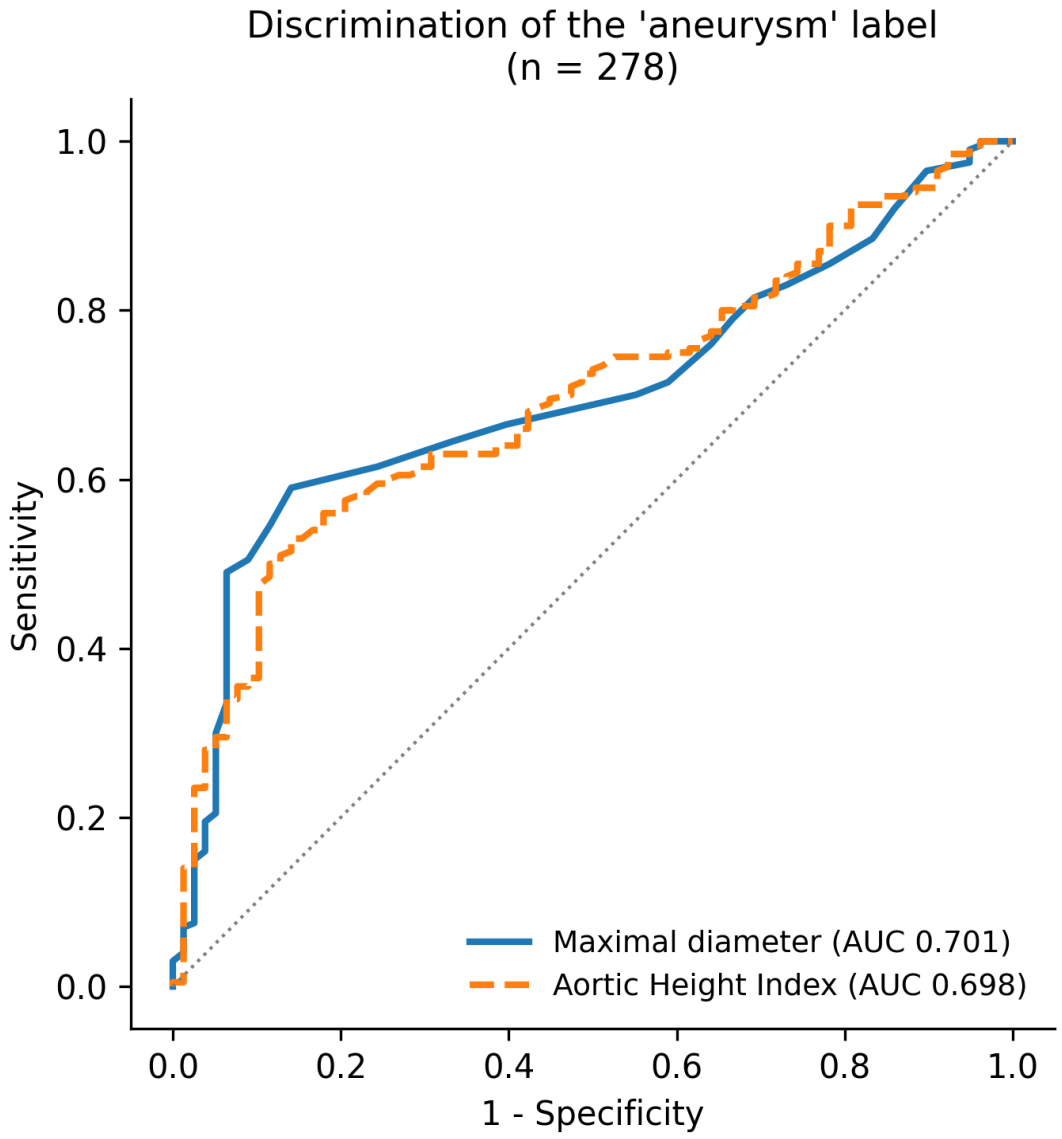
Results/Outcome
Descriptor severity increased monotonically with both diameter (4.06, 4.15, 5.03 cm) and AHI (2.27, 2.41, 2.91 cm/m); $\rho = +0.32$, $p < 0.001$. Discrimination of the "aneurysm" label was moderate and statistically identical for the two metrics (diameter AUC 0.701 vs AHI 0.698; difference +0.004, 95% CI −0.020 to +0.027). Diameter and AHI were near-collinear ($r = 0.971$), as height varied far less than diameter (CV 6.7% vs 29.6%). Of aortas labeled aneurysmal, 72% fell below the AHI surgical threshold; only 6.7% of aortas above it escaped the label.

Conclusion
Descriptor language tracks aortic size directionally but is applied substantially more liberally than guideline thresholds. In an unselected CTA population, AHI provides no discriminative information beyond raw diameter.

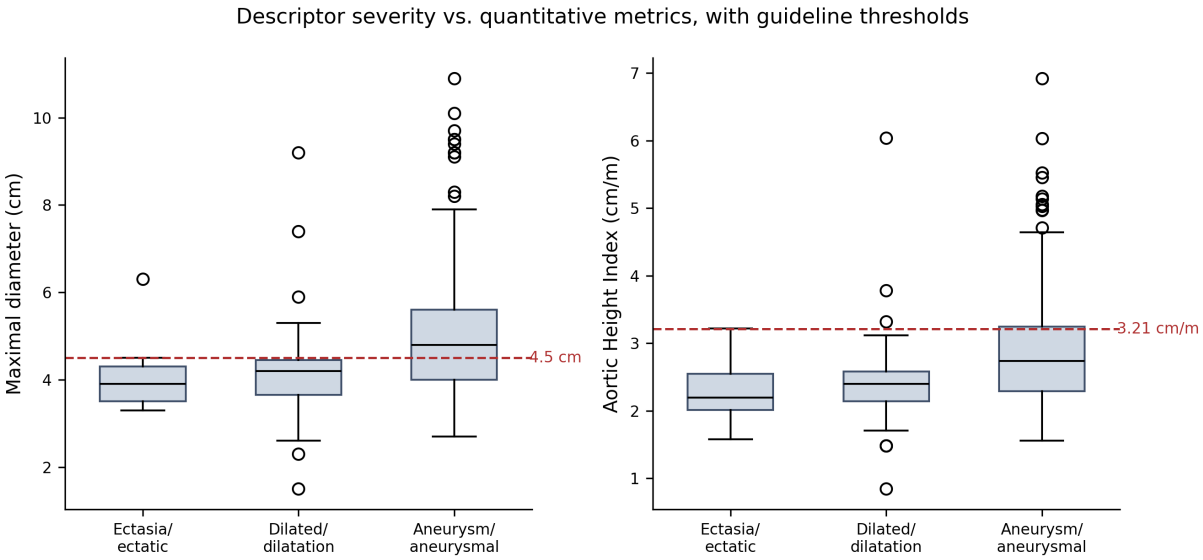
Statement of Impact
LLM extraction makes reporting-language variation auditable at scale; the collinearity of AHI with diameter argues against automated height-indexing as a standalone reporting adjunct outside selected populations.

Descriptor tier	n	Diameter, cm (mean ± SD)	AHI, cm/m (mean ± SD)	≥ 4.5 cm (%)	≥ 5.0 cm (%)	AHI ≥ 3.21 (%)
Ectasia / ectatic	11	4.06 ± 0.85	2.27 ± 0.44	18.2	9.1	9.1
Dilated / dilatation	67	4.15 ± 1.01	2.41 ± 0.62	25.4	6.0	4.5
Aneurysm / aneurysmal	**200**	**5.03 ± 1.52**	**2.91 ± 0.88**	**61.5**	**44.0**	**28.0**

Descriptor tier vs. quantitative metrics and guideline thresholds (n = 278). Spearman ρ = +0.32 for diameter and +0.31 for AHI, both p < 0.001



ROC curves for maximal diameter and Aortic Height Index discriminating the 'aneurysm' descriptor. Bootstrapped AUC difference +0.004 (95% CI −0.020 to +0.027).



Distribution of maximal aortic diameter (left) and Aortic Height Index (right) across the three descriptor severity tiers (n = 278). Boxes show the interquartile range with the median; whiskers extend to 1.5 × IQR, with individual points beyond plotted. Dashed lines mark the ACC/AHA ascending aortic aneurysm threshold (4.5 cm) and the alternate surgical threshold for AHI (3.21 cm/m). Both metrics rise monotonically with descriptor severity (Spearman ρ = +0.32 and +0.31, respectively; both $p < 0.001$), but distributions overlap substantially across tiers, and the majority of aortas labeled aneurysmal fall below both thresholds.

Keywords

large language models; information extraction; aortic aneurysm; aortic height index; structured reporting