nature portfolio

Peer Review File

Deciphering Age- and Sex-Specific Patterns of Coronary Artery Atherosclerosis from a Large Chinese Cohort

Corresponding Author: Professor Dinggang Shen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
1 Although the study samples are drawn from multiple centers across China, it is not clearly stated whether the sample distribution covers both urban and rural areas or regions with different economic levels. Given China's vast size and the significant lifestyle and dietary differences across regions, these factors may affect the results. It is recommended that the authors provide a detailed description of the sample's regional distribution and discuss the applicability of the results in different regions.

2 Although the study emphasizes the advantages of PLASMA in automation, it does not address the practical challenges of its application in clinical workflows, such as computation time and system compatibility issues. It is recommended that the authors include an evaluation of the model's operational efficiency and the feasibility of its clinical deployment to provide a comprehensive assessment of its practical application potential in clinical settings.

3 In the Methods section, the authors should focus on providing a detailed description of the data and the specific implementation of the methods. However, the results of the PLASMA evaluation are placed in this section. It is recommended that the authors organize the model validation content and move it to the Results section to better align with academic writing standards.

4 In the Introduction section, the authors mention that the analysis results are based on the outcomes of PLASMA, but the specific role of PLASMA in the analysis is not clearly highlighted in the Results section. It is recommended that the authors reorganize this part of the writing to clearly demonstrate PLASMA's contribution to the results.

5 The authors mention that the PLASMA results are based on the epoch with the lowest loss on the validation set, which implies that the validation set was involved in the hyperparameter tuning process. This approach may lead to an overestimation of the PLASMA results. It is recommended that the authors designate a separate dataset as an independent test set and ensure that this test set does not participate in the training process of PLASMA, in order to obtain more fair and reliable results.

6 Given the limited amount of data used for training and evaluation of PLASMA, the results may be affected by randomness, potentially leading to an overestimation. To improve the reliability of the results, it is recommended that the authors use five-fold cross-validation to reduce result variability and enhance the model's stability.

7 The authors have divided age into five groups: <45, 45-55, 56-65, 66-75, and >=75, but have not explained the specific rationale behind these groupings. Given that the range of age groups may have a potential impact on the results, it is recommended that the authors further clarify the basis for the groupings. For example, the authors mention that the significant increase in plaque burden in females is observed only after the age of 45-55, but it may actually begin to manifest between 50-55 years, with more gradual changes between 45-50 years. Therefore, a more detailed discussion or adjustment of the age groupings is suggested to improve the accuracy and interpretability of the results.

8 The authors mention that the growth rate for females is close to zero (but less than 0) under the age of 45, and 3.22% per decade for those over 75. However, the authors have divided age into five groups: <45, 45-55, 56-65, 66-75, and >=75. Could you clarify how this result was calculated?

9 The authors mention that 1,237 cases of poor image quality were excluded, but the specific criteria for quality assessment are not clearly defined. It is recommended that the authors provide a more detailed description of the criteria used to determine poor image quality, in order to enhance the transparency and reproducibility of the study.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This is a large scale analysis of consecutive scans from some centres in China. The authors develop a machine learning enabled methodology to perform coronary segmentation, plaque segmentation and labelling. They present the results of the development and validation (internal and external) showing overall good results. Although the scale and methods are appropriate, this is an epidemiology relevant piece of work with results related to the Chinese population.
Specific comments:
1. Amongst the significant limitations of the work is the lack of a true external validation cohort
2. There is limited information on quality control steps that have been taken to ensure that scans are correctly identified and segmented.
3. There seems to be a selection bias, since there has been no randomized community-based screening and invitations for a CT scan for what is essentially an epidemiological piece of work.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The paper can be accepted.

(Remarks on code availability)
The paper can be accepted.

Reviewer #2

(Remarks to the Author)
Thank you for sending me the revised version of the manuscript. This epidemiological piece of work is using state of the art techniques that automate coronary analysis offering interesting insights in CAD phenotyping in a large cohort from China. The authors have adequately addressed the reviewers' comments, providing important supporting data and accordingly altering the manuscript. Limitations are well discussed and the conclusions are appropriate.

(Remarks on code availability)

Responses for Reviewer 1:

1. Although the study samples are drawn from multiple centres across China, it is not clearly stated whether the sample distribution covers both urban and rural areas or regions with different economic levels. Given China's vast size and the significant lifestyle and dietary differences across regions, these factors may affect the results. It is recommended that the authors provide a detailed description of the sample's regional distribution and discuss the applicability of the results in different regions.

   Response:
   We thank the reviewer for this insightful comment and fully agree that economic levels, along with lifestyle and dietary differences associated with China's vast geographic area, may influence the results.

   The cohort in this study was recruited from seven major geographical regions in China — East, North, Northeast, Northwest, Southwest, Central, and South — with sample sizes of each region presented in Figure 1(b). Of the total subjects, 14,923 were from more-developed urban areas and 1,377 were from less-developed rural areas. Details of the regional and urban–rural sample distribution are added to the "Results — Population section" (lines [72–76]) and is summarized in Supplementary Tables 4 and 5.

   To further address the reviewer's concern, we conducted subgroup analyses stratified by urban–rural status and geographical region. These analyses examined differences in overall plaque burden and sex-specific gap in atherosclerosis onset.

   In the rural versus urban analysis, data from third-tier or smaller cities are classified as representing economically less-developed regions, serving as a proxy for rural settings. These participants comprised a minority of the cohort because CCTA is an advanced imaging modality more commonly accessible in urban centers. The observed patterns of atherosclerotic burden and sex-specific progression were highly consistent with the main manuscript findings. Specifically, females in both urban and rural settings demonstrated an approximately 20-year delay in the onset of atherosclerosis compared to males. In females, plaque burden increased in a nonlinear fashion with age, whereas males exhibited a more linear and steady increase (Supplementary Figure 1). Moreover, while rural populations showed a higher total plaque burden, the average total plaque burden in both subgroups remained lower than that observed in North America population (Supplementary Table 4).

   In the analysis across the seven major regions, the overall pattern of delayed onset in females and lower plaque burden relative to North American populations were consistently observed. Most regions showed an approximate 20-year delay, while the Southwestern and Southern regions exhibited longer delays, 24 and 25 years respectively (Supplementary Table 5).

Overall, these supplementary subgroup analyses demonstrate that our main findings are generalizable across Chinese populations with varying economic backgrounds. Although absolute metrics, such as the magnitude of the delayed onset, may differ across regions, the consistent relative differences observed between genders and countries reinforce the applicability of our findings to populations with diverse geographic and socioeconomic backgrounds.

We have incorporated all results and detailed descriptions into Supplementary Information section 1.2, Supplementary Figure 1, and Supplementary Tables 4 and 5, and have also added a discussion of these findings in the manuscript “Discussion” section (lines [244–250] and [268-270]).

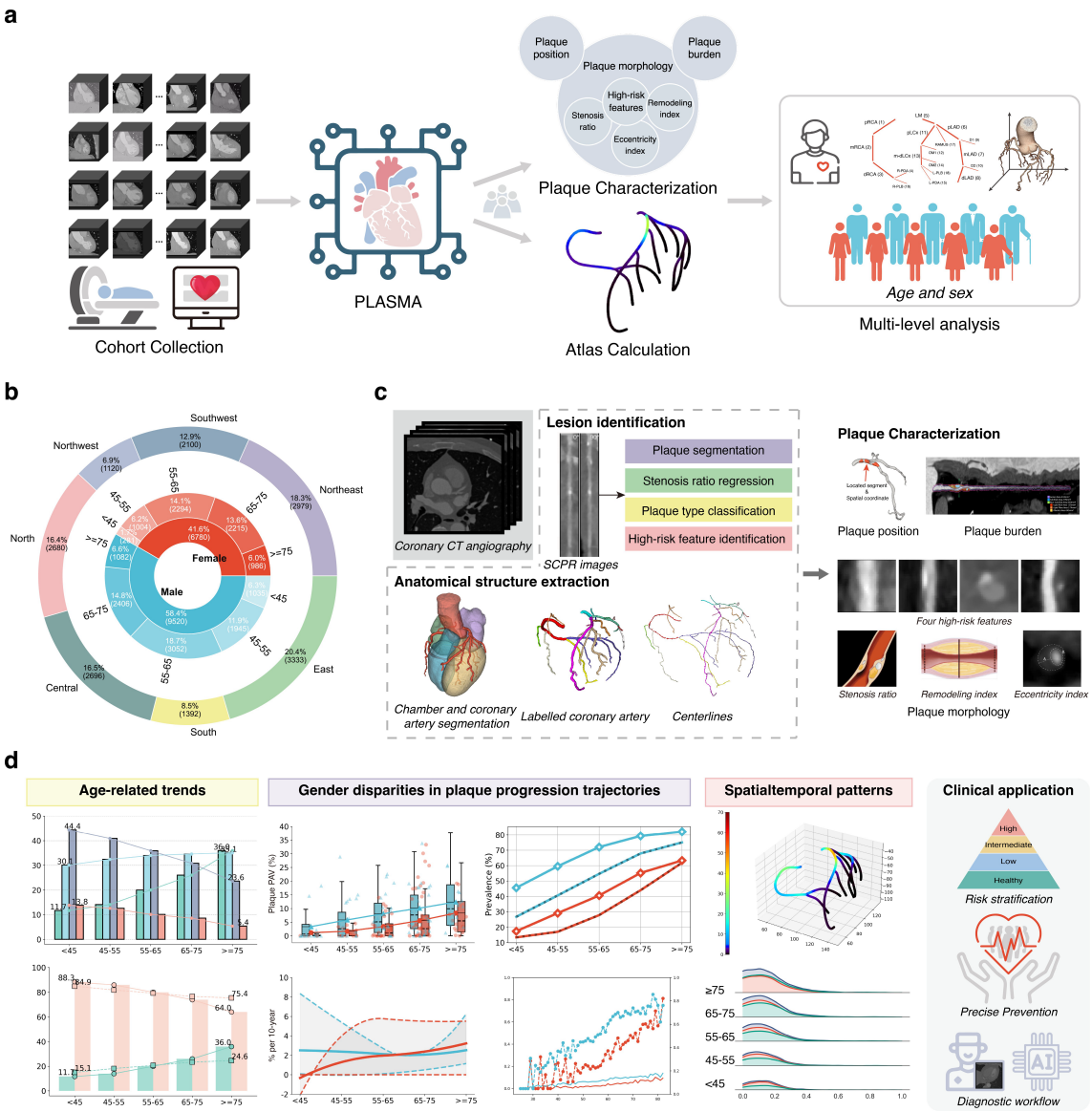
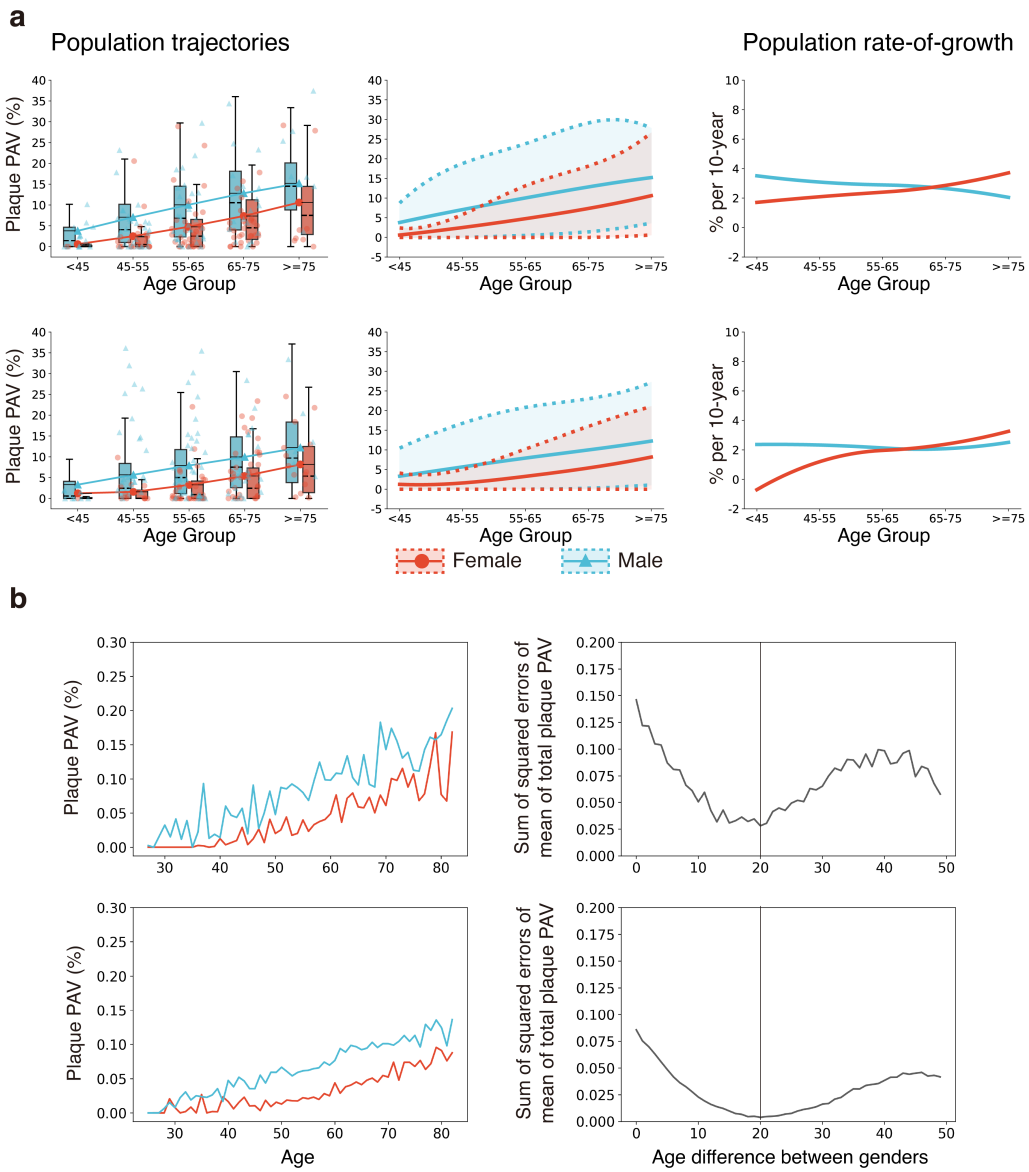


Figure 1. Overview of our study. a, Study approach. A large multi-center CCTA dataset was collected as inputs of the CCTA-based plaque analysis system using machine-learning algorithms (PLASMA) for extracting coronary artery- and plaque-related information. Then, characteristics that are significant for clinical diagnosis were calculated and the plaque distribution atlas were built, based on system outputs. Finally, we conducted analysis at per-subject and per-segment level, and in the continuous space stratified by

age and sex. **b**, Demographic distribution of the study cohort. The multi-layered donut chart illustrates characteristics of the cohort (N = 16,300), stratified by sex (innermost), age (middle), and geographic (outermost) region. **c**, PLASMA-based plaque characterization. Anatomical structures and lesion information were automatically extracted by PLASMA and used to calculate plaque characteristics, including position, burden, and morphology. **d**, Main results and clinical applications.



Supplementary Figure 1: Sex differences in atherosclerosis onset and plaque progression between rural and urban populations. a, Sex differences in plaque burden characteristics of rural (the 1st row) and urban (the 2nd row) populations. **b,** Disparities in atherosclerosis severity between sexes of rural (the 1st row) and urban (the 2nd row) populations.

Supplementary Table 4: Plaque volume characteristics between rural and urban populations.

	Entire Population	Female	Male	p-value
Rural	1377	639	738	
Urban	14923	6141	8782	
Age, y †				
Rural	58.12±11.54 (27-95)	60.53±10.50 (27-95)	56.04±11.99 (27-94)	
Urban	61.80±11.65 (25-97)	63.78±10.66 (26-96)	60.41±12.10 (25-97)	
TP-V, mm^3 #				
Rural	343.88±468.24 165.33 (24.76-469.94)	225.91±337.36 92.27 (0.00-310.35)	446.02±536.95 243.88 (69.53-646.31)	<0.0001
Urban	291.78±415.65 126.16 (0.00-404.29)	177.21±295.72 53.10 (0.00-224.18)	371.90±465.65 207.69 (36.26-535.85)	<0.0001
CP-V, mm^3 #				
Rural	73.86±161.74 5.99 (0.00-69.32)	54.03±118.52 1.76 (0.00-48.24)	91.02±189.85 13.84 (0.00-92.75)	<0.0001
Urban	82.94±165.97 11.26 (0.00-86.30)	58.29±129.53 3.13 (0.00-52.73)	100.17±185.35 19.53 (0.00-112.93)	<0.0001
NCP-V, mm^3 #				
Rural	270.02±353.81 137.67 (21.54-373.42)	171.88±245.12 80.92 (0.00-234.50)	355.00±407.57 222.07 (65.68-504.91)	<0.0001
Urban	208.84±295.60 94.53 (0.00-287.95)	118.92±197.98 39.96 (0.00-156.21)	271.73±333.87 158.89 (28.83-387.58)	<0.0001
FP-V, mm^3 #				
Rural	118.28±174.19 48.65 (4.88-160.18)	81.93±130.49 31.14 (0.00-100.10)	149.75±199.44 70.53 (17.56-210.20)	<0.0001
Urban	98.42±150.11 39.18 (0.00-132.73)	61.06±107.63 17.97 (0.00-74.19)	124.55±168.92 63.66 (10.09-170.46)	<0.0001
FFP-V, mm^3 #				
Rural	123.91±169.19 65.28 (9.40-169.91)	74.21±106.32 35.98 (0.00-98.37)	166.94±199.15 105.75 (31.93-237.17)	<0.0001
Urban	86.03±129.69 36.34 (0.00-115.30)	45.35±80.23 13.45 (0.00-58.32)	114.48±148.71 63.64 (9.79-158.06)	<0.0001
LP-V, mm^3 #				
Rural	27.83±47.70 10.34 (0.22-35.40)	15.74±26.98 4.02 (0.00-19.60)	38.30±58.14 18.40 (2.72-51.01)	<0.0001
Urban	24.39±43.08 6.95 (0.00-30.10)	12.51±26.22 1.52 (0.00-13.31)	32.70±50.05 13.71 (0.67-43.67)	<0.0001

†Data are means±SDs and range (minimum-maximum).
#Data are means±SDs and median (IQR).
TP = total plaque; CP = calcified plaque; NCP = non-calcified plaque; FP = fibrotic plaque; FFP = fibrous fatty plaque; LP = lipid plaque; V = volume; SD = standard deviation; IQR = interquartile range.

Supplementary Table 5: Subgroup analyses across seven major geographical regions in China.

	Entire Population	Female	Male
Sample Size			
East	3333	1461	1872
North	2680	1111	1569
Northeast	2979	1289	1690
Northwest	1120	434	686
Southwest	2100	841	1259
Central	2696	1071	1625
South	1392	573	819
Age, y †			
East	61.71±11.85 (27-96)	63.87±10.94 (28-95)	60.03±12.25 (27-96)
North	60.23±11.18 (26-91)	62.63±10.19 (27-90)	58.53±11.54 (26-91)
Northeast	59.42±11.70 (25-96)	61.17±10.73 (27-94)	58.10±12.22 (25-96)
Northwest	62.31±11.71 (25-97)	64.57±9.94 (31-94)	60.88±12.50 (25-97)
Southwest	62.79±11.52 (28-94)	64.69±10.56 (28-94)	61.52±11.95 (28-94)
Central	62.52±11.63 (25-96)	64.43±10.62 (26-96)	61.27±12.09 (25-93)
South	63.16±11.70 (27-94)	64.88±10.88 (27-94)	61.95±12.11 (28-93)
Total Plaque Volume, mm^3 #			
East	290.43±418.51 131.46 (9.88–393.99)	191.44±309.85 71.49 (0.00–243.14)	367.70±472.63 195.49 (42.20–525.05)
North	375.45±482.78 199.13 (22.49–545.53)	232.61±357.61 83.45 (0.00–293.02)	476.59±531.98 307.13 (76.44–684.98)
Northeast	260.50±398.51 95.62 (0.00–348.25)	147.01±262.37 38.14 (0.00–177.11)	347.07±458.46 168.48 (16.31–493.37)
Northwest	303.97±414.76 142.85 (16.83–425.32)	200.97±324.64 69.21 (0.00–249.87)	369.13±451.02 214.93 (40.39–540.76)
Southwest	255.27±376.37 111.54 (0.00–357.93)	146.59±239.12 42.49 (0.00–193.86)	327.86±430.13 184.46 (29.70–469.77)
Central	296.16±419.91 126.38 (3.89–414.97)	181.48±296.67 53.22 (0.00–233.99)	371.74±469.29 203.96 (34.57–525.55)
South	289.20±391.00 132.48 (9.67–413.59)	174.71±285.10 51.67 (0.00–225.68)	369.30±433.05 218.64 (49.88–559.47)
Delayed Years, y			
East		18	
North		21	
Northeast		19	
Northwest		20	
Southwest		24	
Central		20	
South		25	

†Data are means±SDs and range (minimum-maximum).
#Data are means±SDs and median (IQR).
SD = standard deviation; IQR = interquartile range.

2. Although the study emphasizes the advantages of PLASMA in automation, it does not address the practical challenges of its application in clinical workflows, such as computation time and system compatibility issues. It is recommended that the authors include an evaluation of the model's operational efficiency and the feasibility of its clinical deployment to provide a comprehensive assessment of its practical application potential in clinical settings.

Response:

We thank the reviewer for this insightful comment regarding the practical aspects of clinical deployment. In response, we have incorporated a detailed evaluation of the system's operational efficiency and feasibility of clinical deployment.

a. Evaluation of operational efficiency

As suggested, we systematically evaluated the computation time of PLASMA, defined as the time elapsed from the input of a CCTA scan into the PLASMA system to the output of interpretation results, including coronary segmentation, plaque detection and quantification, stenosis grading, high-risk feature classification, and post-processing. Across different hardware configurations, PLASMA achieved mean computation times ranging from approximately 17.12 seconds (NVIDIA GeForce RTX 4060 Ti with Intel® Core™ i9-12900K) to 29.95 seconds (NVIDIA RTX 2080 Ti with Intel® Xeon® Silver 4110), as detailed in Supplementary Table 2. Notably, even on mid-range GPUs, the system consistently delivered results in under one minute per case. RAM usage remained below ~6 GB, and GPU memory consumption was under ~5 GB across all setups.

Supplementary Table 2: Computation time and memory use of PLASMA across five GPUs. Results are based on testing with a single GPU. Mean per-case computation time was measured on 232 CCTA scans processed through the fully automated pipeline, including coronary segmentation, plaque detection and quantification, stenosis grading and high-risk feature classification.

GPU+CPU Model	Computation Time, s	RAM, GB	GPU, GB
NVIDIA GeForce RTX 4060 Ti + Intel® Core™ i9-12900K	17.12±3.72 (10.12–26.00)	3.78±0.35 (3.34–5.47)	3.99±0.21 (3.49–4.60)
NVIDIA GeForce RTX 3060 + Intel® Core™ i9-10900X	19.56±5.54 (9.87–32.46)	3.79±0.33 (3.36–5.51)	3.91±0.32 (2.81–4.68)
NVIDIA RTX A4000 + Intel® Xeon® Silver 4210R	25.25±6.16 (13.32–36.37)	3.79±0.35 (3.32–5.72)	
Quadro RTX 4000 + Intel® Xeon® Silver 4210	26.00±6.80 (14.61–41.37)	2.76±0.36 (2.32–4.79)	
NVIDIA GeForce RTX 2080 Ti + Intel® Xeon® Silver 4110	29.95±7.31 (16.71–45.96)	2.77±0.34 (2.31–4.59)	

Values are presented as mean±standard deviation (SD) and range (minimum–maximum) for computation time, RAM usage, and GPU memory usage. The reported RAM consumption reflects the parallel resource requirements during GPU-accelerated computations.

According to our prior study (Yang, W. *et al., La Radiol Med.* 2023 Feb;128(3):307-315), radiologist post-processing and interpretation for anatomical and stenosis assessment averaged ~187 seconds per case using a dedicated workstation. Furthermore, Lin et al. reported that manual plaque quantification using semi-automated research software

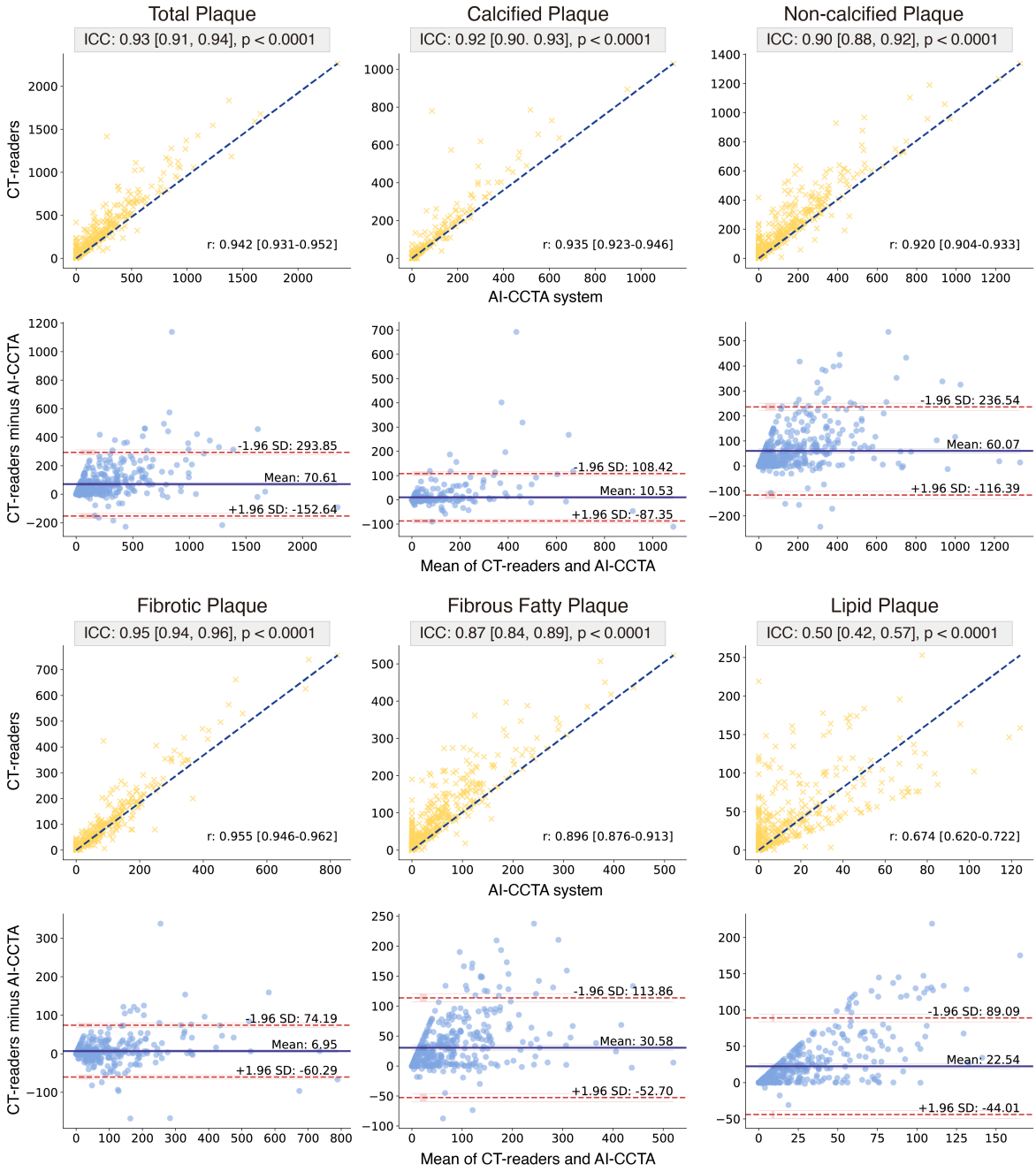
required approximately 26 minutes per patient (Lin, A. *et al., The Lancet Digit. Heal.* 2022 April;4(4):e256–e265).

These comparisons demonstrate that PLASMA offers substantial gains in operational efficiency over conventional manual or semi-automated workflows, reducing the end-to-end analysis time to well below one minute per case on mid-range hardware, thereby facilitating easier integration into clinical workflows. The evaluation results and their interpretation regarding system operational efficiency have been incorporated into the Supplementary Information.

b. Feasibility of clinical deployment

Our prior multi-center, multi-reader study (Yang, W. *et al., La Radiol Med.* 2023 Feb;128(3):307-315) validated the clinical usability of PLASMA for stenosis diagnosis. This study included 374 CCTA cases from five clinical sites and 12 radiologists, using invasive coronary angiography as the reference standard, and demonstrated that the DL-based system effectively reduced post-processing and diagnostic time, while improving diagnostic accuracy — particularly benefiting less experienced readers. Another recent study (Yu, L. *et al., La Radiol Med.* 2024 Aug;129(8):1173-1183) showed that following the deployment of the coronary segmentation software of PLASMA in a tertiary hospital, both the number of CCTA scans performed daily and the number of reconstructions completed per radiologist per day approximately has doubled due to its high efficiency since January, 2021.

Building on these validated components, the current study expands PLASMA's capabilities with the integration of an automated plaque analysis module. The added external validation on an independent dataset (n = 439) showed a high level of agreement with expert annotations, with ICC values of 0.93 for total plaque volume, 0.92 for calcified plaque, and 0.90 for non-calcified plaque; Pearson's r values ranged from 0.92 to 0.94. The results of external validation have been added in "Results — Performance Evaluation of PLASMA" section (lines [200–205]) and demonstrated in Extended Figure 7.



Extended Data Figure 7. External validation of PLASMA against CT readers for plaque volume measurements (*n* = 439). Scatter plots (1st and 3rd rows) show Pearson's correlation coefficients, while Bland–Altman plots (2nd and 4th rows) illustrate the agreement between PLASMA and expert CT readers across six plaque types: total, calcified, non-calcified, fibrotic, fibrous-fatty, and lipid. Intraclass correlation coefficients (ICC), along with their 95% confidence intervals and significance values, are displayed in light gray boxes. ICC: intraclass correlation coefficient; SD: standard deviation.

Taken together, the system demonstrated diagnostic reliability, combined with its efficient computation time (under one minute per case) and modest hardware demands (<6 GB RAM, <5 GB GPU), supports its feasibility for routine clinical use. PLASMA provides an end-to-end automated solution that is both time-efficient and scalable, making it adaptable to various clinical settings.

3. In the “Methods” section, the authors should focus on providing a detailed description of the data and the specific implementation of the methods. However, the results of the PLASMA evaluation are placed in this section. It is recommended that the authors organize the model validation content and move it to the “Results” section to better align with academic writing standards.

   Response:
   We thank the reviewer for this helpful suggestion regarding the organization of the manuscript. In the revised manuscript, all internal and external validation outcomes, including performance metrics for coronary segmentation, plaque detection and segmentation, stenosis grading, and high-risk plaque feature classification, have now been relocated to the “Results — Performance Evaluation of PLASMA” section (lines [157–211]). The “Methods — AI models: PLASMA” section now focuses on detailed descriptions of model architectures and implementation, training procedures, and validation design.

4. In the Introduction section, the authors mention that the analysis results are based on the outcomes of PLASMA, but the specific role of PLASMA in the analysis is not clearly highlighted in the Results section. It is recommended that the authors reorganize this part of the writing to clearly demonstrate PLASMA’s contribution to the results.

   Response:
   We thank the reviewer for this valuable comment. To address this point, we added descriptions at the beginning of the “Results — Performance Evaluation of PLASMA” section (lines [158-166]) to explicitly clarify PLASMA’s role in the study. As illustrated in Figure 1(a) and (c), PLASMA served as the core analytical tool, providing fully automated CCTA interpretation, including cardiac and coronary anatomy analysis, plaque detection and quantification, stenosis grading, and high-risk feature classification. All automated segmentations were further reviewed by trained technicians with manual corrections applied when needed. From these outputs, we derived the comprehensive plaque characteristics that underpin the population analyses presented in this work. These revisions ensure that PLASMA’s contribution to the study’s analytical workflow is clearly and transparently conveyed.

5. The authors mention that the PLASMA results are based on the epoch with the lowest loss on the validation set, which implies that the validation set was involved in the hyper-parameter tuning process. This approach may lead to an overestimation of the PLASMA results. It is recommended that the authors designate a separate dataset as an independent test set and ensure that this test set does not participate in the training process of PLASMA, in order to obtain more fair and reliable results.

   Response:

We sincerely thank the reviewer for this important comment. To address the concern regarding potential overestimation due to validation-based model selection, we collected an independent dataset ($n = 439$) from three medical centers that was not involved in model training or hyper-parameter tuning and performed an external validation of the plaque segmentation module.

As presented in Extended Data Figure 7 and detailed in "Results — Performance Evaluation of PLASMA" section (lines [200-205]), the model demonstrated good agreement with expert annotations, achieving an ICC of 0.93 (95% CI: 0.91–0.94) and Pearson's r of 0.94 (95% CI: 0.93–0.95) for total plaque volume. Similar strong consistency was observed for calcified (ICC: 0.92, Pearson's r: 0.94) and non-calcified plaques (ICC: 0.90, Pearson's r: 0.92). The results of external validation are in agreement with those from five-fold cross validation on the internal dataset, as outlined in the next section.

These independent results confirm the generalizability and robustness of the model beyond the internal dataset, addressing the reviewer's concerns.

6. Given the limited amount of data used for training and evaluation of PLASMA, the results may be affected by randomness, potentially leading to an overestimation. To improve the reliability of the results, it is recommended that the authors use five-fold cross-validation to reduce result variability and enhance the model's stability.

Response:

We thank the reviewer for this valuable suggestion. In response, we conducted a five-fold cross-validation on the plaque segmentation module using 1,560 cases from the internal dataset. The results demonstrated excellent consistency across all folds. For total plaque volume, the mean ± standard deviation (SD) of ICC was 0.92 ± 0.0274, and that of Pearson's r was 0.93 ± 0.0295. Similar high consistency was observed for calcified and non-calcified plaque volumes. Detailed fold-wise results have been descripted in "Results — Performance Evaluation of PLASMA" section (lines [195-199]) and summarized in Supplementary Table 1.

Supplementary Table 1: Performance of PLASMA versus experts plaque volume quantification in internal five-fold cross-validation setting (n=1,560). Intraclass correlation coefficients (ICC) and Pearson's correlation coefficients were computed between PLASMA and expert assessments for total, calcified, non-calcified, and lipid plaque volumes.

Plaque Type	Fold	ICC (95% CI)	p-value	Pearson's r (95% CI)
Total Plaque Volume	Fold0	0.94 (0.93–0.95)	<0.0001	0.95 (0.93–0.96)
	Fold1	0.89 (0.86–0.91)		0.90 (0.87–0.92)
	Fold2	0.89 (0.86–0.91)		0.90 (0.87–0.92)
	Fold3	0.94 (0.93–0.95)		0.96 (0.95–0.97)
	Fold4	0.94 (0.93–0.96)		0.95 (0.94–0.96)
	avg.	**0.92 ± 0.0274**	<0.0001	**0.93 ± 0.0295**
Calcified Plaque Volume	Fold0	0.99 (0.98–0.99)	<0.0001	0.99 (0.99–0.99)
	Fold1	0.98 (0.97–0.98)		0.99 (0.98–0.99)
	Fold2	0.99 (0.99–0.99)		1.00 (0.99–1.00)
	Fold3	0.99 (0.99–1.00)		1.00 (0.99–1.00)
	Fold4	0.99 (0.99–0.99)		1.00 (0.99–1.00)
	avg.	**0.99 ± 0.0045**	**<0.0001**	**1.00 ± 0.0055**
Non-calcified Plaque Volume	Fold0	0.90 (0.88–0.92)	<0.0001	0.91 (0.89–0.93)
	Fold1	0.83 (0.79–0.87)		0.84 (0.81–0.87)
	Fold2	0.85 (0.81–0.88)		0.86 (0.82–0.88)
	Fold3	0.91 (0.88–0.92)		0.94 (0.93–0.95)
	Fold4	0.92 (0.90–0.94)		0.93 (0.91–0.94)
	avg.	**0.88 ± 0.0396**	**<0.0001**	**0.90 ± 0.0439**
Lipid Plaque Volume	Fold0	0.71 (0.65–0.77)	<0.0001	0.77 (0.72–0.81)
	Fold1	0.59 (0.51–0.66)		0.64 (0.57–0.70)
	Fold2	0.55 (0.47–0.63)		0.65 (0.58–0.71)
	Fold3	0.58 (0.50–0.65)		0.73 (0.67–0.78)
	Fold4	0.70 (0.63–0.75)		0.79 (0.74–0.83)
	avg.	**0.63 ± 0.0737**	**<0.0001**	**0.72 ± 0.0684**

ICC = Intraclass Correlation Coefficient; CI = Confidence Interval. The avg. represent the mean ± standard deviation across folds. Higher ICC and Pearson's r values indicate better agreement.

For the remaining key functional modules, both internal and external validation results have been previously reported, including coronary artery segmentation (Li, M. *et al., Radiology.* 2022 Oct;306(3):e221393, Yu, L. *et al., La Radiol Med.* 2024 Aug;129(8):1173-1183), as well as plaque detection and stenosis grading (Yang, W. *et al., La Radiol Med.* 2023 Feb;128(3):307-315). These prior validations provide strong support for the overall reliability of PLASMA outputs.

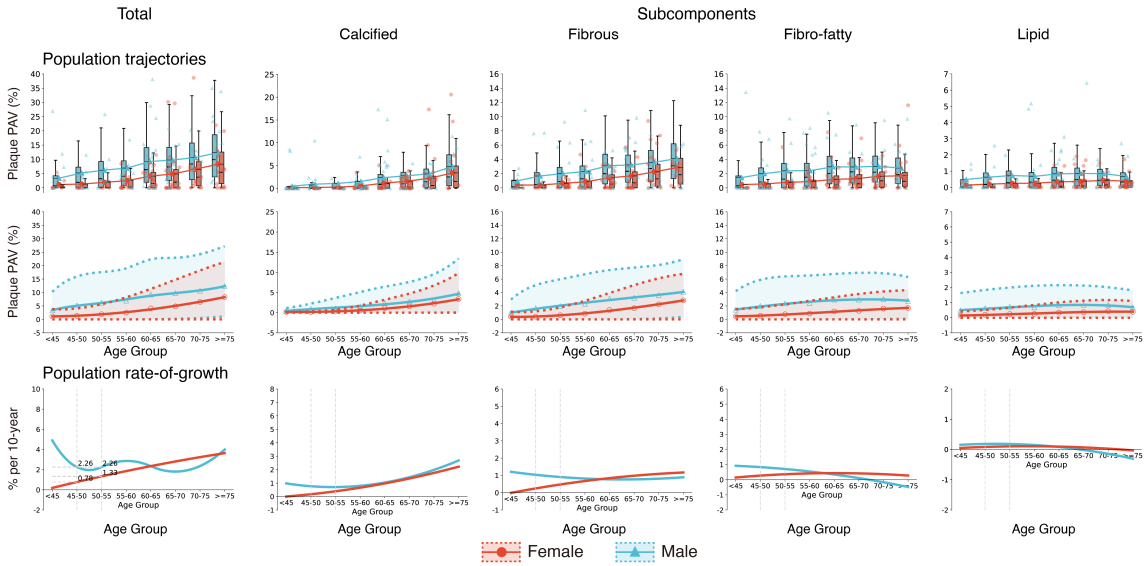
7. The authors have divided age into five groups: <45, 45-55, 56-65, 66-75, and >=75, but have not explained the specific rationale behind these groupings. Given that the range of age groups may have a potential impact on the results, it is recommended that the authors further clarify the basis for the groupings. For example, the authors mention that the significant increase in plaque burden in females is observed only after the age of 45-55, but it may actually begin to manifest between 50-55 years, with more gradual changes between 45-50 years. Therefore, a more detailed discussion or adjustment of the age groupings is suggested to improve the accuracy and interpretability of the results.

Response:

We thank the reviewer for this thoughtful and constructive suggestion. The initial age groupings (<45, 45–55, 56–65, 66–75, ≥75 years) were based on prior literature (Tzimas, G. *et al. JACC:Cardiovasc. Imaging,* 2024 Feb;17(2):165–175) and selected to balance the age and sex distribution in our dataset, ensuring sufficient statistical power for each group.

We fully agree with the reviewer that the 45–55 age range, particularly in females, represents a period of significant hormonal changes, potentially influencing plaque progression differently within narrower age spans. In response, we conducted an additional 5-year stratified analysis. The results show that plaque burden begins to increase gradually between 45–50 years but accelerates markedly in the 50–55 age group, aligning with the reviewer's insight. Specifically, we observed a modest 0.2% increase in total plaque burden (PAV) from <45 to 45–50 years, followed by a sharper 0.6% rise from 45–50 to 50–55 years, with sustained elevations thereafter. The estimated plaque burden growth rate increased from 0.78% to 1.33% (increase in PAV per decade) between 45–50 and 50–55 group intervals, as shown in Extended Data Figure 3. Here, "per decade" reflects the unit used in our fitting model, where each step represents a 10-year age interval. If absolute age values (e.g., [40, 50, 60, 70, 80]) were used instead — corresponding to a 1-year interval — the slope magnitude would be numerically smaller by a factor of 10 (e.g., 0.322% per year instead of 3.22% per decade), while the trend remains unchanged.

We have updated the relevant sections in the "Results — Sex Disparities in Atherosclerosis Onset and Plaque Progression" (lines [100–104]) and "Discussion" (lines [251–257]), and detailed results are illustrated in Extended Data Figure 3.



Extended Data Figure 3. Sex differences in plaque burden characteristics using finer age stratification. (The 1st and 2nd rows) Box plots of plaque PAV distributions across age groups, with PAV-age trajectories of mean (dotted lines indicating 10% and 90% centiles) estimated by cubic B-spline (smooth factor=0.1). (The 3rd row) Rates of PAV change over the lifespan, derived from PAV-age trajectories. All plots are stratified by sex (female, red; male, blue).

8. The authors mention that the growth rate for females is close to zero (but less than 0) under the age of 45, and 3.22% per decade for those over 75. However, the authors have divided age into five groups: <45, 45-55, 56-65, 66-75, and >=75. Could you clarify how this result was calculated?

Response:

Thank you for this thoughtful question.

To improve the clarity, we revised the description in the "Methods — Statistical Analysis" section (lines [495–502]) to explain how these trajectories were calculated.

As noted in "Methods — Statistical Analysis" section, we applied cubic B-spline interpolation to fit a continuous age-related curve of plaque burden (PAV), allowing us to capture the underlying trend across the lifespan. The growth rate trajectory was then calculated as the first derivative of the interpolated curve.

Specifically, the near-zero (slightly negative) growth rate under age 45 and the 3.22% per decade growth rate over age 75 were derived by extracting the values of the mean PAV growth rate trajectory for females at these two age groups (<45 and ≥75 years). It is important to note that the slightly negative value observed under age 45 does not reflect a biologically meaningful decrease, but results from the mathematical behavior of the spline model, which balances curve smoothness across all age intervals.

9. The authors mention that 1,237 cases of poor image quality were excluded, but the specific criteria for quality assessment are not clearly defined. It is recommended that the authors provide a more detailed description of the criteria used to determine poor image quality, in order to enhance the transparency and reproducibility of the study.

Response:

We thank the reviewer for this valuable comment. We agree that a clear description of the image quality assessment criteria is essential for transparency and reproducibility. To address this, we have added a detailed description of the image quality evaluation in the "Methods — Cohort" section (lines [345–349]). Specifically, we described an established 4-point scale scoring criteria (Dai, X. *et al., Eur Radiol.* 2019 Aug;29(8):4349–4356) used for evaluating CCTA image quality — excellent (3), good (2), sufficient (1), poor (0). Cases with scores of 0 or 1 were excluded, and diagnostic readability of all four proximal coronary segments (RCA, LM, LAD, LCx) was required for quantitative plaque assessment. We believe this clarification enhances the transparency and reproducibility of the study.

Responses for Reviewer 2:

1. Amongst the significant limitations of the work is the lack of a true external validation cohort.

Response:

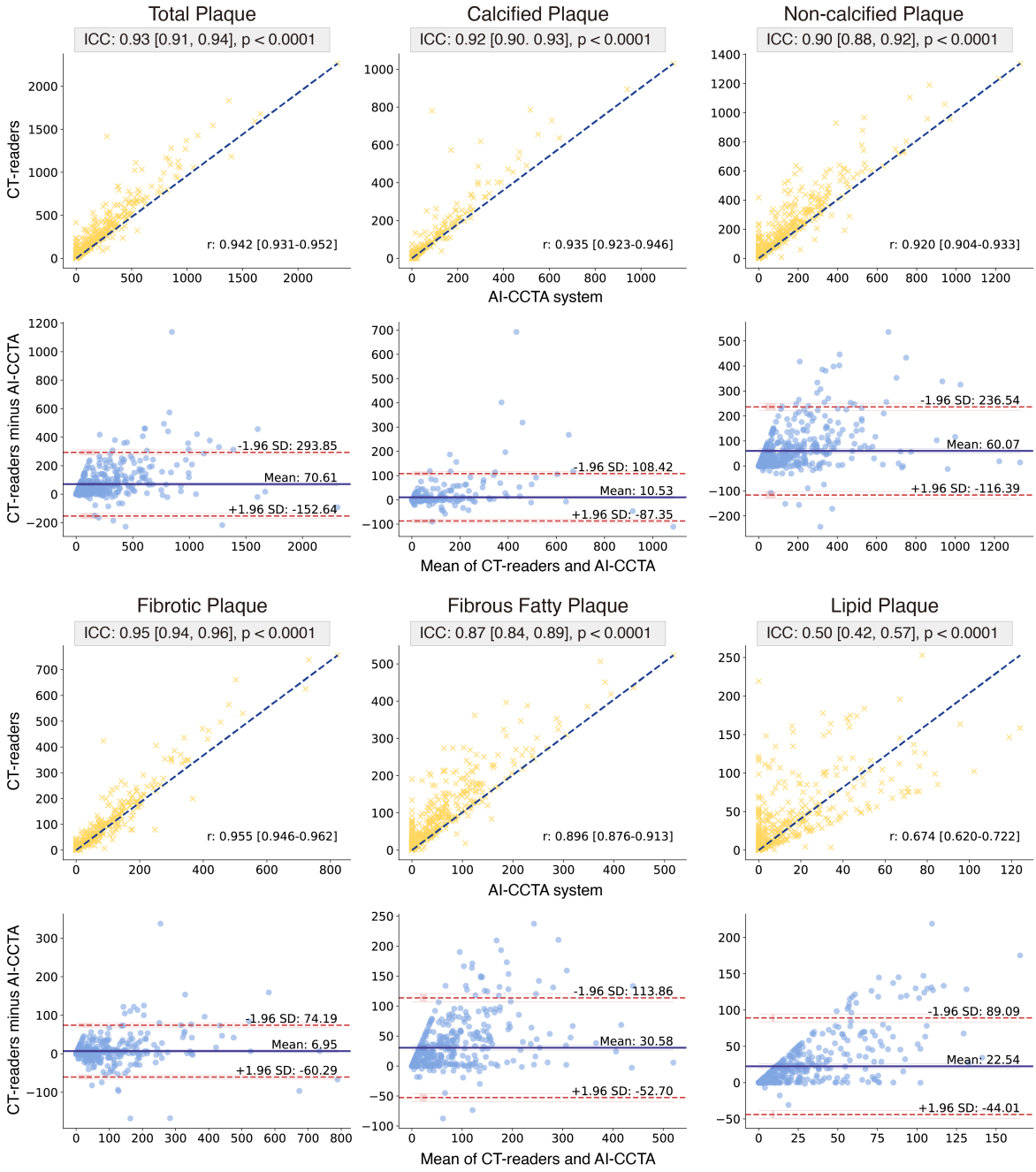
We thank the reviewer for this important comment. To address this concern, we have conducted external validations on the PLASMA modules that directly impact our study outcomes.

As described in the revised "Methods — AI models: PLASMA" section (lines [414–419] and [451–463]), these validations were performed using independent datasets collected from multiple medical centers, with diverse scanner vendors and real-world imaging protocols, ensuring robust technical performance of the system across varied clinical settings.

Specifically, the coronary segmentation and labeling module was externally validated using two independent datasets from two medical centers, covering four major CT vendors and a broad range of vascular conditions — including native vessels, CTO, stents, bypass grafts, and origin anomalies. Full results were published in our prior studies (Li, M. *et al., Radiology.* 2022 Oct;306(3):e221393, Yu, L. *et al., La Radiol Med.* 2024 Aug;129(8):1173-1183).

For the plaque detection and stenosis grading module, a multi-reader, multi-center study was previously conducted (Yang, W. *et al., La Radiol Med.* 2023 Feb;128(3):307-315), involving 374 cases from five external hospitals across two geographic regions in China.

In this revision, we further conducted external validation of the plaque segmentation module, using a newly collected independent dataset (n = 439) from three centers in Northern and Eastern China. The results demonstrated strong agreement with expert annotations and are presented in "Results — Performance Evaluation of PLASMA" section (lines [200–205]) and Extended Data Figure 7. In parallel, five-fold cross-validation was conducted on the internal dataset (n = 1,560), with the results summarized in Supplementary Table 1, showing comparable performance to the external validation.



Extended Data Figure 7. External validation of PLASMA against CT readers for plaque volume measurements (*n* = 439). Scatter plots (1st and 3rd rows) show Pearson's correlation coefficients, while Bland–Altman plots (2nd and 4th rows) illustrate the agreement between PLASMA and expert CT readers across six plaque types: total, calcified, non-calcified, fibrotic, fibrous-fatty, and lipid. Intraclass correlation coefficients (ICC), along with their 95% confidence intervals and significance values, are displayed in light gray boxes. ICC: intraclass correlation coefficient; SD: standard deviation.

Supplementary Table 1: Performance of PLASMA versus experts plaque volume quantification in internal five-fold cross-validation setting (n=1,560). Intraclass correlation coefficients (ICC) and Pearson's correlation coefficients were computed between PLASMA and expert assessments for total, calcified, non-calcified, and lipid plaque volumes.

Plaque Type	Fold	ICC (95% CI)	p-value	Pearson's r (95% CI)
Total Plaque Volume	Fold0	0.94 (0.93–0.95)		0.95 (0.93–0.96)
	Fold1	0.89 (0.86–0.91)		0.90 (0.87–0.92)
	Fold2	0.89 (0.86–0.91)	<0.0001	0.90 (0.87–0.92)
	Fold3	0.94 (0.93–0.95)		0.96 (0.95–0.97)
	Fold4	0.94 (0.93–0.96)		0.95 (0.94–0.96)
	avg.	**0.92 ± 0.0274**	<0.0001	**0.93 ± 0.0295**
Calcified Plaque Volume	Fold0	0.99 (0.98–0.99)		0.99 (0.99–0.99)
	Fold1	0.98 (0.97–0.98)		0.99 (0.98–0.99)
	Fold2	0.99 (0.99–0.99)	<0.0001	1.00 (0.99–1.00)
	Fold3	0.99 (0.99–1.00)		1.00 (0.99–1.00)
	Fold4	0.99 (0.99–0.99)		1.00 (0.99–1.00)
	avg.	**0.99 ± 0.0045**	**<0.0001**	**1.00 ± 0.0055**
Non-calcified Plaque Volume	Fold0	0.90 (0.88–0.92)		0.91 (0.89–0.93)
	Fold1	0.83 (0.79–0.87)		0.84 (0.81–0.87)
	Fold2	0.85 (0.81–0.88)	<0.0001	0.86 (0.82–0.88)
	Fold3	0.91 (0.88–0.92)		0.94 (0.93–0.95)
	Fold4	0.92 (0.90–0.94)		0.93 (0.91–0.94)
	avg.	**0.88 ± 0.0396**	**<0.0001**	**0.90 ± 0.0439**
Lipid Plaque Volume	Fold0	0.71 (0.65–0.77)		0.77 (0.72–0.81)
	Fold1	0.59 (0.51–0.66)		0.64 (0.57–0.70)
	Fold2	0.55 (0.47–0.63)	<0.0001	0.65 (0.58–0.71)
	Fold3	0.58 (0.50–0.65)		0.73 (0.67–0.78)
	Fold4	0.70 (0.63–0.75)		0.79 (0.74–0.83)
	avg.	**0.63 ± 0.0737**	**<0.0001**	**0.72 ± 0.0684**

ICC = Intraclass Correlation Coefficient; CI = Confidence Interval. The avg. represent the mean ± standard deviation across folds. Higher ICC and Pearson's r values indicate better agreement.

2. There is limited information on quality control steps that have been taken to ensure that scans are correctly identified and segmented.

Response:

We thank the reviewer for raising this important point.

To ensure that CCTA scans were correctly identified and segmented, we implemented a rigorous quality control process throughout the cohort selection and analysis workflow.

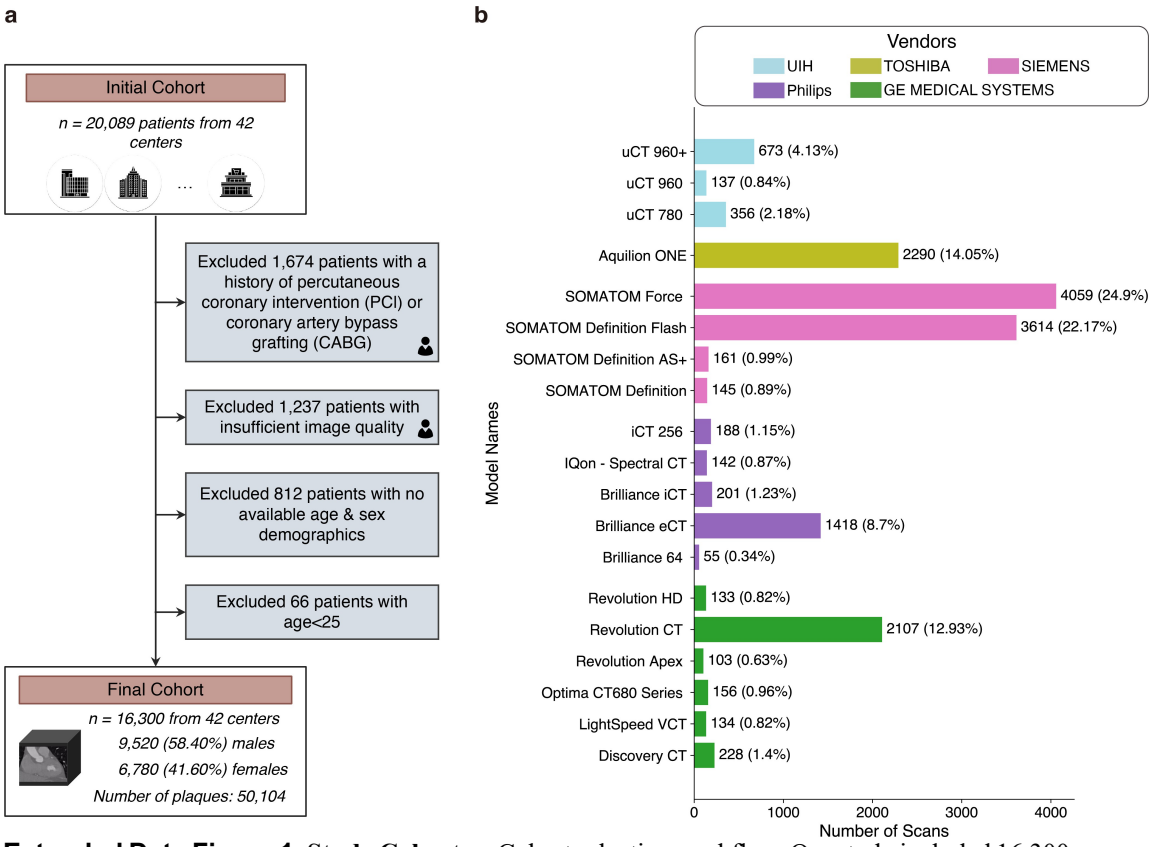
During cohort selection, as clarified in the revised "Methods — Cohort" section (lines [343–352]) and illustrated in the cohort selection workflow (Extended Data Figure 1(a)), exclusion criteria were applied under the supervision of certified CT analysts:

(1) Scans with poor image quality were excluded based on evaluation by certified CT analysts using a standardized 4-point scale, in accordance with established criteria (Dai, X. *et al., Eur Radiol.* 2019 Aug;29(8):4349–4356). Specifically, scans rated as 0 (severe artifact, non-diagnostic) or 1 (moderate artifact, low diagnostic confidence) were classified as poor quality and excluded from the study. In addition, cases were only included if all four major proximal coronary segments (RCA, LM, LAD, and LCx) were adequately visualized to ensure complete coronary assessment.

(2) Patients with a history of cardiac interventions, including percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG), were excluded. These cases

were initially identified by a task-specific AI model that detected the presence of stents or grafts in the CCTA images, and were further confirmed through manual review.

In the automated analysis pipeline, extensive internal and external validation core PLASMA modules were conducted to assess the reliability of automatic identification and segmentation, as described in the "Results — Performance Evaluation of PLASMA" section (lines [167–211]). Eight trained technicians were invited to review the automated segmentation results, and made necessary manual editions when segmentation showed suboptimal results.



Extended Data Figure 1. Study Cohort. a, Cohort selection workflow. Our study included 16,300 subjects from 42 centers in mainland China. **b**, CT scanner vendor distribution in our study.

3. There seems to be a selection bias, since there has been no randomized community-based screening and invitations for a CT scan for what is essentially an epidemiological piece of work.

Response:

We appreciate the reviewer's insightful comment.

As acknowledged in the revised "Discussion" section (lines [299–304]), the study cohort was derived from clinically indicated CCTA scans rather than from a randomized,

community-based screening population, which introduces a degree of selection bias toward symptomatic or at-risk individuals.

Nonetheless, this real-world clinical cohort reflects the current diagnostic landscape in China, where CCTA is primarily performed in symptomatic or at-risk populations. Interpreted within this clinical context, our findings offer valuable epidemiological insights into sex- and age-specific patterns of atherosclerosis from multiple levels, which are highly relevant for patient risk stratification and healthcare decision-making. Future studies incorporating randomized community-based screening are needed to enhance the generalizability of our findings.

We have further clarified this limitation in the revised “Discussion” section and strengthened the context of symptomatic populations.