



Artificial intelligence denoising in cardiac photon-counting CT: mitigating BMI-related noise degradation while maintaining clinical interchangeability

Lisa Urban, Josephine Berger, Jan M. Brendel, Judith Herrmann, Patrick Krumm, Konstantin Nikolaou, Saif Afat, Andreas Brendlin ^{*} 

Department of Diagnostic and Interventional Radiology, Eberhard-Karls University, Hoppe-Seyler-Str. 3, 72076 Tuebingen, Germany

ARTICLE INFO

Keywords:

Tomography
X-ray computed
Photon counting
Coronary angiography
Artificial intelligence
Image enhancement

ABSTRACT

Objectives: To evaluate whether an artificial intelligence-based denoising (AID) algorithm can effectively attenuate body mass index (BMI)-related noise degradation in cardiac photon-counting detector CT (PCD-CT) while maintaining strict quantitative diagnostic equivalence and clinical interchangeability compared with quantum iterative reconstruction (QIR).

Materials and Methods: This retrospective study included 100 patients undergoing non-contrast coronary artery calcium (CAC) scoring and contrast-enhanced coronary CT angiography (CCTA) on a dual-source PCD-CT. Images were reconstructed using standard QIR and a PCD-CT-weighted AID algorithm. Subjective image quality was assessed by four blinded raters. Objective metrics included noise magnitude (noise power spectrum) and spatial resolution (edge rise distance). Frequentist generalized estimating equations (GEE) evaluated reconstruction effects and BMI interactions. Clinical interchangeability (Agatston scoring, coronary age, CAD-RADS, and plaque characterization) was assessed using Bayesian regions of practical equivalence (ROPE).

Results: AID significantly improved subjective image quality and reduced overall noise magnitude by approximately 64 % in CAC and 78 % in CCTA ($p < 0.001$). GEE models revealed that AID robustly attenuated BMI-driven noise escalation by 44 % (CAC) and 78 % (CCTA, $p < 0.001$ for interactions in both). Noise texture was stable in CCTA datasets, but revealed smoothening in CAC datasets. However, spatial resolution was preserved and remained stable across varying patient habitus. ROPE analysis proved strict diagnostic equivalence between QIR and AID, yielding near-certain probabilities (> 0.999) of practical equivalence for continuous Agatston scores, categorical CAD-RADS severity classification, and specific coronary plaque composition.

Conclusions: The PCD-CT-weighted AID algorithm improves subjective and objective image quality in both CAC and CCTA, while maintaining diagnostic equivalence and workflow efficiency compared to QIR.

1. Introduction

Coronary artery disease is a leading cause of death worldwide [1]. Arteriosclerotic narrowing or occlusion of coronary arteries reduces myocardial blood flow, manifesting clinically as angina, myocardial infarction, or heart failure [2]. Early risk assessment and detection are

essential to prevent such events [3]. Cardiac CT is well suited for excluding coronary heart disease in patients with low to intermediate pre-test probability [2]. Coronary CT angiography (CCTA) and coronary artery calcium (CAC) scoring are established, non-invasive methods [4]. CAC quantifies coronary calcifications, while CCTA detects calcified and non-calcified plaques, enabling reliable exclusion or confirmation of

Abbreviations: AID, Artificial Intelligence-based denoising Algorithms; BMI, Body Mass Index; BPM, Beats per Minute; CAC, Coronary Artery Calcium; CAD-RADS, Coronary Artery Disease-Reporting and Data System; CCTA, Coronary Computed Tomography Angiography; CT, Computed Tomography; CTDI_{vol}, Computed Tomography Dose Index; DLP, Dose-length Product; EID, Energy Integrating Detectors; ERD, Edge Rise Distance; ESF, Edge Spread Function; FDR, False Discovery Rate; GEE, Generalized Estimating Equations; HDI, Highest Density Interval; HU, Hounsfield Units; IRB, Institutional Review Board; MCMC, Markov Chain Monte Carlo; MRI, Magnetic Resonance Imaging; NPS, Noise Power Spectrum; PCD-CT, Photon Counting Detector Computed Tomography; QIR, Quantitative Iterative Reconstruction; ROI, Region of Interest; ROPE, Region of practical equivalence; SD, Standard Deviation; SSDE, Size-specific Dose Estimate.

^{*} Corresponding author.

E-mail address: andreas.brendlin@med.uni-tuebingen.de (A. Brendlin).

<https://doi.org/10.1016/j.ejrad.2026.112921>

Received 28 October 2025; Received in revised form 10 March 2026; Accepted 4 May 2026

Available online 5 May 2026

0720-048X/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

coronary stenosis [5]. Photon-counting detector CT (PCD-CT) improves cardiovascular imaging with higher intrinsic spatial resolution and reduced electronic noise; however, overall image quality remains fundamentally limited by photon starvation and residual noise, particularly in patients with a high body mass index (BMI) [6]. Artificial intelligence-based denoising (AID) has recently shown promise in mitigating such severe noise penalties and enhancing contrast across various imaging applications [7,8]. However, its efficacy in specifically addressing BMI-driven noise escalation in cardiac PCD-CT requires rigorous clarification. Furthermore, because PCD-CT detectors differ fundamentally in noise texture and behavior from traditional energy-integrating detectors (EID-CT), the diagnostic safety of applying EID-CT AID solutions in this novel ecosystem remains uncertain. We therefore investigated the impact of a PCD-CT weighted AID algorithm on subjective and objective image quality, with a specific focus on its robustness across varying patient body habitus in both CAC and CCTA, compared to standard quantum iterative reconstruction (QIR). We hypothesized that PCD-CT AID would effectively attenuate BMI-related noise degradation while strictly maintaining quantitative diagnostic equivalence and clinical interchangeability.

2. Materials and methods

2.1. Study design and study population

This retrospective single-center study was approved by the local institutional review board (IRB) with a waiver of informed consent (#167/2022BO2) and conducted in accordance with the Declaration of Helsinki (1964, revised 2024). An a priori sample size calculation was performed in G*Power (version 3.1.9.7, Franz Faul, University of Kiel, Germany) using an F-test framework for repeated-measures ANOVA (within factors) as a pragmatic approximation for correlated within-subject comparisons [9]. Assuming a small effect size of $f = 0.148$, $\alpha = 0.05$, a power = 0.95, a correlation of 0.5 among repeated measures, and a nonsphericity correction of 1, the required total sample size was 100. All CT scans acquired with the same scanner between August 1, 2022, and August 1, 2023, were screened for eligibility. In the first step, patients with non-cardiac CT examinations were excluded. In the second step, only the most recent 100 patients undergoing CAC or CCTA were included, while earlier cases were excluded as beyond IRB approval.

For each patient, demographic and anthropometric data (age, sex, height, weight), presenting symptoms and cardiovascular risk factors were extracted from clinical records at the time of examination. Body mass index (BMI, kg/m^2) was calculated from recorded height and weight. Radiation dose parameters, including CTDIvol (mGy), dose-length product (DLP, $\text{mGy}\cdot\text{cm}$), and size-specific dose estimate (SSDE, mGy), were extracted using the commercial dose-management software DoseM® (Infinit Europe GmbH, Frankfurt am Main, Germany).

2.2. Image acquisition and reconstruction parameters

All imaging was performed on a first-generation dual-source PCD-CT scanner (NAEOTOM Alpha; Siemens Healthineers). Patients were scanned supine, head-first, with arms elevated. Following resting electrocardiogram and echocardiogram assessments, intravenous metoprolol (for resting heart rates >60 bpm) and sublingual nitroglycerin were administered if no contraindications were present. Patients first underwent an ECG-gated non-contrast scan for CAC scoring. This was followed by contrast-enhanced CCTA using Ultravist 370 (Bayer Vital) and a dual-head power injector (ulrichINJECT CT motion), with the specific acquisition mode adapted to the patient's heart rate. All images were reconstructed using Quantum Iterative Reconstruction (QIR; Siemens Healthineers). The resulting images were subsequently post-processed using a vendor-neutral artificial intelligence-based denoising algorithm adapted to PCD-CT noise characteristics (ClariCT.AI vPCD 1.0, ClariPi Inc.). Both reconstructions were generated from identical raw

datasets, with the AI algorithm applied as a post-processing step to the QIR images. Consequently, the comparison in this study reflects QIR reconstruction versus QIR followed by AI-based denoising (QIR + AID). For readability, the latter is abbreviated as “AID” throughout the manuscript. Detailed scanning, contrast administration, and reconstruction parameters are summarized in [Supplementary Tables 1 and 2](#).

2.3. Subjective image quality comparison

Subjective image quality was assessed using a blinded, forced-choice paired-comparison design. An independent researcher removed identifying information and prepared all dataset permutations. For each patient, CAC and CCTA datasets reconstructed with both QIR and AID were included. To eliminate spatial display bias and assess intra-rater reliability, all possible pair combinations and left/right arrangements were generated. This included cross-comparisons between different reconstructions (e.g., QIR vs. AID, AID vs. QIR) and internal control comparisons of identical reconstructions (e.g., QIR vs. QIR, AID vs. AID).

Four radiologists (with 2, 5, 8, and 10 years of cardiac CT experience) independently evaluated the randomized, blinded combinations using MicroDicom (version 2024.1, Sofia, Bulgaria). For each trial, a reference dataset was displayed on the left and a rating dataset on the right. Raters assessed perceived image noise, contrast differences, and structural sharpness across mediastinal, abdominal, and bone windows, with manual window adjustments permitted.

Raters assigned a score to the rating (right) image relative to the reference (left) image: -1 if the rating dataset was inferior, 0 if equal, and $+1$ if superior relative to the reference. Because each comparison appeared in both spatial orientations, directional scores were sign-harmonized and aggregated to yield a net reconstruction preference score (bounded between -1 and $+1$) for each patient and dataset. Ratings from the identical image pairs were isolated to calculate intra-rater reliability metrics.

2.4. Objective image quality comparison

2.4.1. Noise power spectrum (NPS) computation

Two-dimensional noise power spectra (NPS) were computed from homogeneous liver regions using IMQUEST (version 7.3.1.0; Duke University Health System, Durham, NC, USA). These were radially averaged to yield one-dimensional NPS curves as a function of spatial frequency (f , mm^{-1}), with NPS amplitude expressed in $\text{HU}^2 \cdot \text{mm}^2$.

Noise magnitude was quantified using the area under the NPS curve (AUC_{NPS}). Noise texture was evaluated via the peak spatial frequency (Peak_{NPS}), defined as the frequency at which the radially averaged NPS maximized. To minimize discretization bias, this peak location was refined using local quadratic interpolation around the maximal frequency bin. The empirical frequency sampling resolution (Δf) was approximately 0.0157mm^{-1} . [Fig. 1](#) illustrates representative ROI placements within homogeneous liver regions and the corresponding radially averaged NPS curves for the CAC and CCTA datasets.

2.4.2. Spatial resolution

Spatial resolution and structural delineability were evaluated via edge response analysis using Fiji (version 1.54p; NIH, USA), specifically measuring Edge Rise Distance (ERD) and edge slope. To ensure clinical relevance and account for tissue heterogeneity, we analyzed the sternum-to-parasternal tissue interface for the CAC protocol, and the contrasted atrium-to-myocardium interface for the CCTA protocol. Rectangular regions of interest (ROIs) were placed at identical spatial coordinates on both QIR and AID reconstructions to guarantee anatomical alignment. Hounsfield Unit (HU) values were averaged per pixel column to generate one-dimensional line profiles ([Fig. 2](#)). Raw line profiles underwent standardized preprocessing, which included removing missing values, averaging duplicate spatial coordinates,

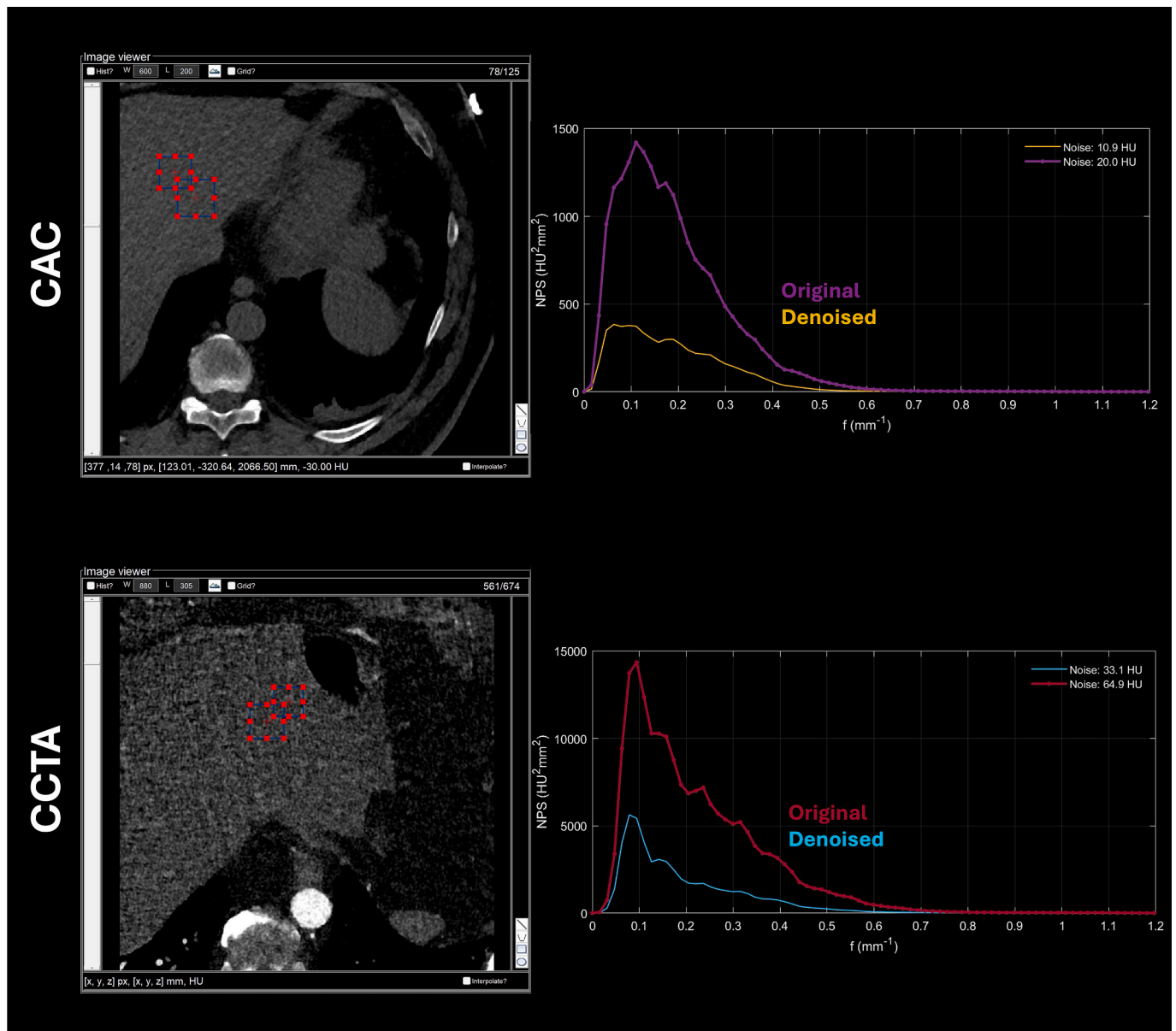


Fig. 1. Representative CAC and CCTA images illustrating ROI placement for NPS analysis. The red markers indicate the homogeneous liver regions used for noise sampling. Corresponding radially averaged NPS curves demonstrate reduced noise magnitude in denoised images compared with the original reconstructions.

applying a five-point moving average filter, and automatically orienting the profiles to ensure a monotonically rising edge. The empirical edge spread function (ESF) was fitted to a four-parameter logistic function using bounded non-linear least squares optimization:

$$f(x) = a + \frac{b}{1 + \exp\left(-\frac{x-x_0}{s}\right)}$$

where a is the baseline intensity, b the step height, x_0 the inflection point, and s the slope scale. ERD was calculated as the physical distance between the 10% and 90% intensity thresholds of the fitted step height, ensuring accurate physical scaling by incorporating dataset pixel spacing. The maximum edge slope was analytically derived at the inflection point as $b/(4s)$ and converted to HU/mm . Finally, to compute aggregated mean ESFs without artificially broadening structural boundaries, individual curves were edge-centered by shifting x_0 to zero, interpolated to a common spatial grid, and averaged.

2.5. Clinical interchangeability

To assess clinical interchangeability, both the QIR and AID series were evaluated sequentially per patient for the CAC and CCTA datasets:

2.5.1. CAC analysis

Datasets were loaded into the CT Cardiac workflow of the multi-modal diagnostic software syngo.via (version VB60H; Siemens Healthineers, Forchheim, Germany) for automated characterization of coronary arteries and calcifications. Manual corrections were applied when necessary. To compare the impact of the reconstructions on the clinical workflow, the Agatston score and coronary age were recorded both before and after manual correction.

2.5.2. CCTA analysis

Datasets were processed using Careverse AI Digital Heart (version 1.0; CAREVERSE TECHNOLOGY PTE. LTD., Singapore) to extract Coronary Artery Disease-Reporting and Data System (CAD-RADS) scores and plaque characteristics. All AI-generated outputs were subsequently

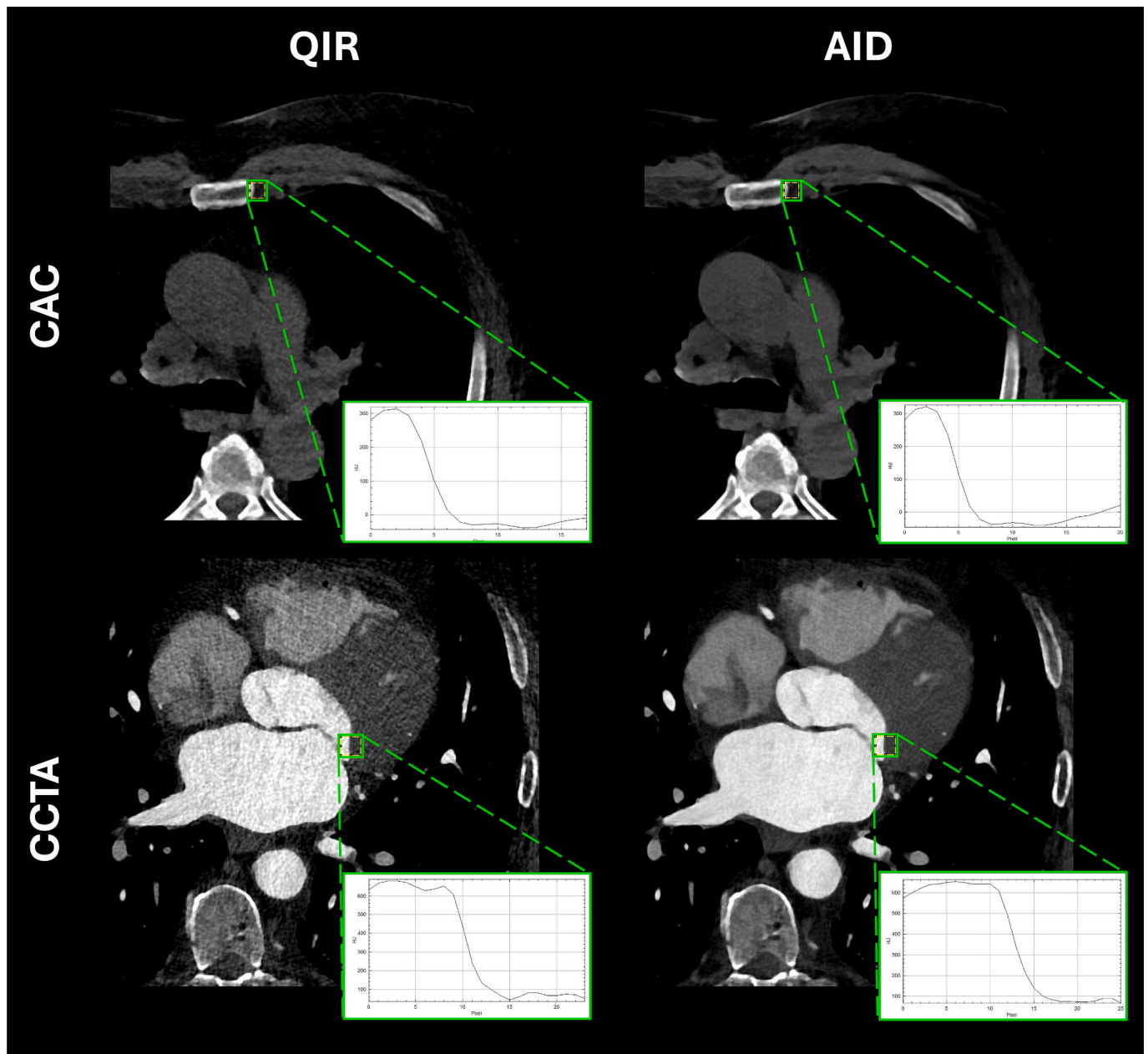


Fig. 2. Representative CAC and CCTA images illustrating ROI placement for ERD analysis (QIR left, AID right). Enlarged regions show the analyzed tissue interfaces and the corresponding one-dimensional edge spread profiles derived from column-averaged Hounsfield Unit values.

reviewed and verified by a certified senior cardiac radiologist with 15 years of experience.

2.6. Statistical analysis

Although the study was powered via an F-test approximation, final inferential analyses employed Generalized Estimating Equations (GEEs) and Bayesian hierarchical models to more rigorously account for within-patient and within-vessel data dependencies, and to formally evaluate practical equivalence.

All statistical analyses were pre-specified and conducted separately for the CAC and CCTA datasets. Continuous variables are reported as mean $\pm$ standard deviation (SD). All frequentist tests were two-sided. To control for multiple comparisons, p-values derived from GEEs were pooled and adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) procedure, with statistical significance defined as an FDR-

corrected $p < 0.05$.

2.6.1. Image quality assessment

Intra- and inter-rater agreement for ordinal subjective image quality ratings ($-1, 0, 1$) was evaluated using Spearman's rank correlation. Following verification of sufficient agreement, ratings were symmetrically harmonized and aggregated at the patient level. To evaluate reconstruction effects and potential interactions with body mass index (BMI), frequentist GEEs (accounting for patient-level clustering) were applied to these aggregated subjective scores, as well as to objective metrics of noise magnitude (AUC_NPS) and spatial resolution (ERD and edge slope).

2.6.2. Diagnostic interchangeability and practical equivalence

For endpoints where the primary objective was to demonstrate practical similarity rather than superiority, Bayesian inference with a

Region of Practical Equivalence (ROPE) was employed. Paired relative differences in continuous physical metrics (noise texture and relative edge sharpness) and clinical continuous data (Agatston scores) were evaluated using robust Bayesian Student-t models. Diagnostic equivalence of categorical and ordinal clinical endpoints (e.g., coronary age, CAD-RADS scores, plaque character) was evaluated using Bayesian latent mixed-effects models. These models mapped categorical outcomes to an underlying continuous latent scale while incorporating random intercepts for patient and coronary vessel to account for hierarchical anatomical clustering.

2.6.3. Software

All analyses and Bayesian modeling were performed in Python (version ≥ 3.10) utilizing statsmodels for frequentist inference and the NumPyro probabilistic programming framework (JAX backend) for Bayesian Markov Chain Monte Carlo (MCMC) sampling. Extended details regarding GEE model specifications, Bayesian link functions, and ROPE thresholds are provided in the [Supplementary Material](#).

3. Results

3.1. Study population

The initial database search yielded 8,643 CT examinations performed between August 1, 2022, and August 1, 2023, on the same first-

generation dual-source PCD-CT scanner. After applying predefined exclusion criteria, 100 patients were enrolled (mean age 57 ± 12 years; mean BMI 27 ± 4 ; 26 female). [Fig. 3](#) illustrates the study design, including patient selection, image reconstruction, and analysis workflow. Detailed demographic, anthropometric, and clinical characteristics are summarized in [Table 1](#).

3.2. Subjective image quality comparison

3.2.1. Inter- and intra-rater agreement

Strong inter-rater agreement was observed across all comparisons (Spearman $R = 0.763$, 95%CI 0.663–0.836, $p < 0.001$), supporting the validity of pooling ratings into aggregated dataset-level preference scores. Intra-rater reliability analyses demonstrated similarly strong consistency across repeated evaluations; see [Supplementary Table 2](#) for rater-specific details.

3.2.2. Subjective image quality scores

Higher mean subjective scores were observed for AID compared with QIR in both protocols. In the CAC dataset, mean aggregated scores were 0.139 ± 0.098 for AID and -0.138 ± 0.100 for QIR. In the CCTA dataset, corresponding values were 0.131 ± 0.106 for AID and -0.133 ± 0.106 for QIR. These descriptive results indicate a consistent directional preference for AID prior to regression modeling.

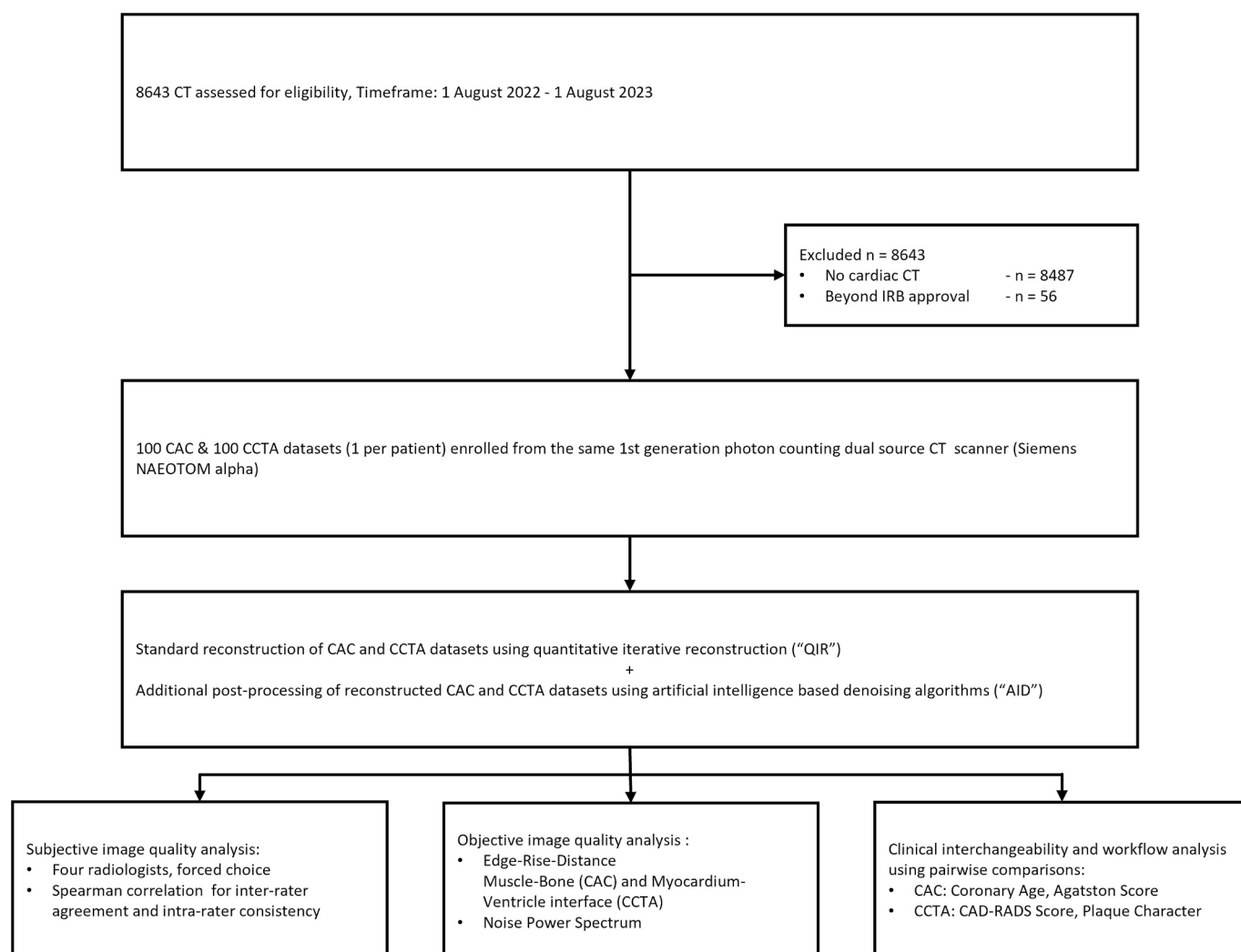


Fig. 3. Study flowchart.

Table 1
Patient DEMOGRAPHICS AND CLINICAL CHARACTERISTICS.

Characteristic	P1 (F)			P2 (SS)			P3 (R)			Total		
	Female	Male	Overall	Female	Male	Overall	Female	Male	Overall	Female	Male	Overall
Population (n)	1	21	22	25	49	74	n/a	4	4	26	74	100
Age (years)	80 ± 0	54 ± 13	55 ± 14	59 ± 11	58 ± 12	58 ± 11	n/a	55 ± 9	55 ± 9	60 ± 12	57 ± 12	57 ± 12
Weight (kg)	75 ± 0	89 ± 13	88 ± 13	75 ± 15	86 ± 18	82 ± 17	n/a	104 ± 18	104 ± 18	75 ± 15	87 ± 17	84 ± 17
Height (cm)	160 ± 0	181 ± 8	180 ± 9	166 ± 6	180 ± 6	175 ± 9	n/a	180 ± 4	180 ± 4	166 ± 6	180 ± 7	177 ± 9
BMI (kg/m ²)	29 ± 0	27 ± 3	27 ± 3	27 ± 5	26 ± 4	27 ± 5	n/a	32 ± 5	32 ± 5	27 ± 5	27 ± 4	27 ± 4
Symptoms												
Dyspnea	0 (0%)	2 (9.5%)	2 (9.1%)	7 (28.0%)	5 (10.2%)	12 (16.2%)	n/a	1 (25.0%)	1 (25.0%)	7 (26.9%)	8 (10.8%)	15 (15.0%)
Angina pectoris	0 (0%)	2 (9.5%)	2 (9.1%)	1 (4.0%)	3 (6.1%)	4 (5.4%)	n/a	0 (0%)	0 (0%)	1 (3.8%)	5 (6.8%)	6 (6.0%)
Syncope	0 (0%)	1 (4.8%)	1 (4.5%)	0 (0%)	4 (8.2%)	4 (5.4%)	n/a	0 (0%)	0 (0%)	0 (0%)	5 (6.8%)	5 (5.0%)
Vertigo	0 (0%)	1 (4.8%)	1 (4.5%)	4 (16.0%)	3 (6.1%)	7 (9.5%)	n/a	0 (0%)	0 (0%)	4 (15.4%)	4 (5.4%)	8 (8.0%)
Chest pain	0 (0%)	2 (9.5%)	2 (9.1%)	4 (16.0%)	9 (18.4%)	13 (17.6%)	n/a	1 (25.0%)	1 (25.0%)	4 (15.4%)	12 (16.2%)	16 (16.0%)
Palpitations	0 (0%)	1 (4.8%)	1 (4.5%)	2 (8.0%)	1 (2.0%)	3 (4.1%)	n/a	0 (0%)	0 (0%)	2 (7.7%)	2 (2.7%)	4 (4.0%)
CVRF												
Arterial hypertension	0 (0%)	7 (33.3%)	7 (31.8%)	7 (28.0%)	13 (26.5%)	20 (27.0%)	n/a	3 (75.0%)	3 (75.0%)	7 (26.9%)	23 (31.1%)	30 (30.0%)
Dyslipidemia	0 (0%)	8 (38.1%)	8 (36.4%)	10 (40.0%)	14 (28.6%)	24 (32.4%)	n/a	2 (50.0%)	2 (50.0%)	10 (38.5%)	24 (32.4%)	34 (34.0%)
Diabetes	0 (0%)	1 (4.8%)	1 (4.5%)	2 (8.0%)	2 (4.1%)	4 (5.4%)	n/a	0 (0%)	0 (0%)	2 (7.7%)	3 (4.1%)	5 (5.0%)
Smoking	0 (0%)	2 (9.5%)	2 (9.1%)	4 (16.0%)	11 (22.4%)	15 (20.3%)	n/a	3 (75.0%)	3 (75.0%)	4 (15.4%)	16 (21.6%)	20 (20.0%)
Adiposity	0 (0%)	4 (19.0%)	4 (18.2%)	6 (24.0%)	5 (10.2%)	11 (14.9%)	n/a	2 (50.0%)	2 (50.0%)	6 (23.1%)	11 (14.9%)	17 (17.0%)
Familial predisposition	0 (0%)	3 (14.3%)	3 (13.6%)	6 (24.0%)	7 (14.3%)	13 (17.6%)	n/a	2 (50.0%)	2 (50.0%)	6 (23.1%)	12 (16.2%)	18 (18.0%)

Table 1: Continuous data are presented as mean ± standard deviation (SD). Categorical data are presented as counts (n) and valid percentages (%). % = Protocol 1 (Ultra-fast flash); % = Protocol 2 (Step-and-shoot); % = Protocol 3 (Retrospective ECG-gated helical); % = Cardiovascular Risk Factors; % = Body Mass Index; % = not applicable / no patients in this subgroup.

3.2.3. CAC datasets

In the non-contrast CAC datasets, AID yielded a significantly higher net preference score compared to QIR at the cohort's mean BMI ($\beta = 0.276$, standard error [SE] 0.0195, $p < 0.001$, 95% confidence interval [CI] 0.238 to 0.314). This modeled effect size closely aligns with the observed difference in raw mean scores (0.139 vs − 0.138), reflecting a consistent patient-level preference for AID rather than isolated extreme

ratings. Furthermore, patient body habitus did not significantly influence subjective image quality under QIR ($\beta = -0.001$, SE 0.002; $p = 0.781$), and there was no significant interaction between BMI and the reconstruction algorithm ($\beta = 0.000$, SE 0.004; $p = 0.952$). Consequently, reader preference for AID remained robust and stable across varying patient sizes.

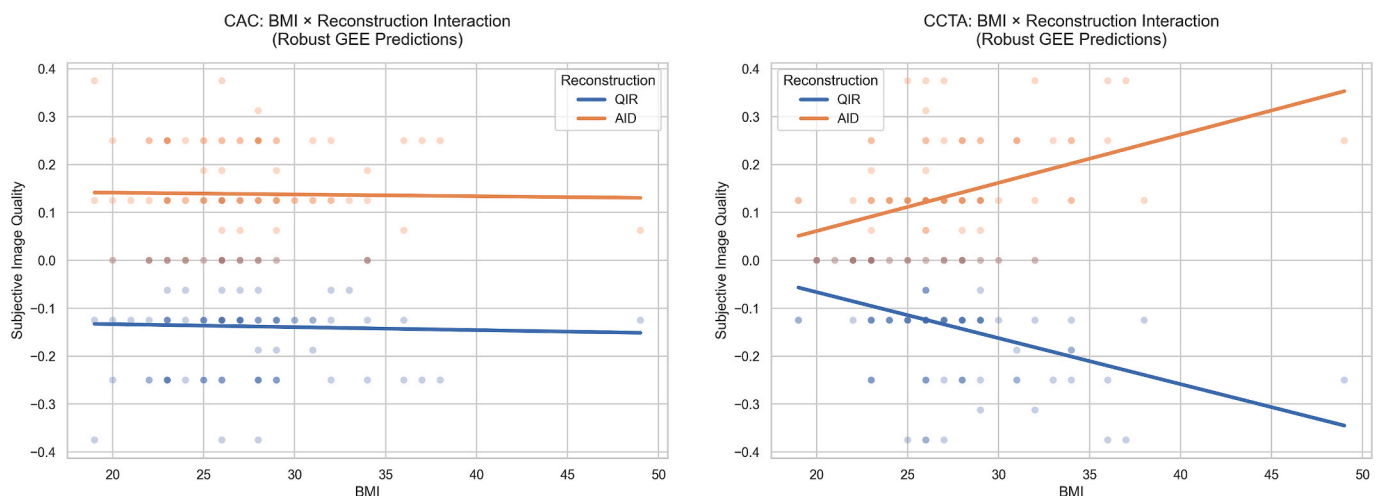


Fig. 4. Modeled effect of BMI on subjective image quality for QIR and AID reconstructions. GEE marginal effect plots for the CAC (left) and CCTA (right) protocols. The y-axis represents the aggregated subjective image quality preference score (bounded from −1 to +1), and the x-axis represents patient BMI. Lines indicate the estimated marginal means, and shaded bands represent 95 % confidence intervals. Left: In the non-contrast CAC protocol, AID demonstrates consistent superiority over QIR regardless of patient habitus, with parallel trendlines indicating no significant interaction ($p = 0.952$). Right: In the contrast-enhanced CCTA protocol, subjective quality under QIR declines significantly with increasing BMI, whereas AID effectively mitigates this degradation, resulting in a significant positive interaction ($p < 0.001$).

3.2.4. CCTA datasets

Similarly, in the contrast-enhanced CCTA datasets, AID demonstrated a strong subjective preference over QIR ($\beta = 0.264$, SE 0.019, 95 % CI: 0.227 to 0.302; $p < 0.001$). This modeled effect size again corresponds closely to the observed mean difference between the reconstructions (0.131 vs. -0.133). Notably, while higher BMI was associated with a significant decline in perceived image quality under QIR ($\beta = -0.010$, SE 0.002, 95 % CI: -0.014 to -0.006; $p < 0.001$), a significant positive interaction was observed between AID and BMI ($\beta = 0.020$, SE 0.004, 95 % CI: 0.011 to 0.028; $p < 0.001$). This indicates that AID effectively mitigated BMI-related subjective image degradation,

maintaining superior visual clarity even in larger patients. Fig. 4 displays the modeled effect of BMI on subjective image quality for QIR and AID reconstructions. Fig. 5 shows representative CAC and CCTA reconstructions, highlighting the improved clarity with AID. The corresponding numerical values are listed in Table 2.

3.3. Objective image quality comparison

3.3.1. Noise magnitude

3.3.1.1. CAC datasets. In the non-contrast CAC protocol, at mean BMI,

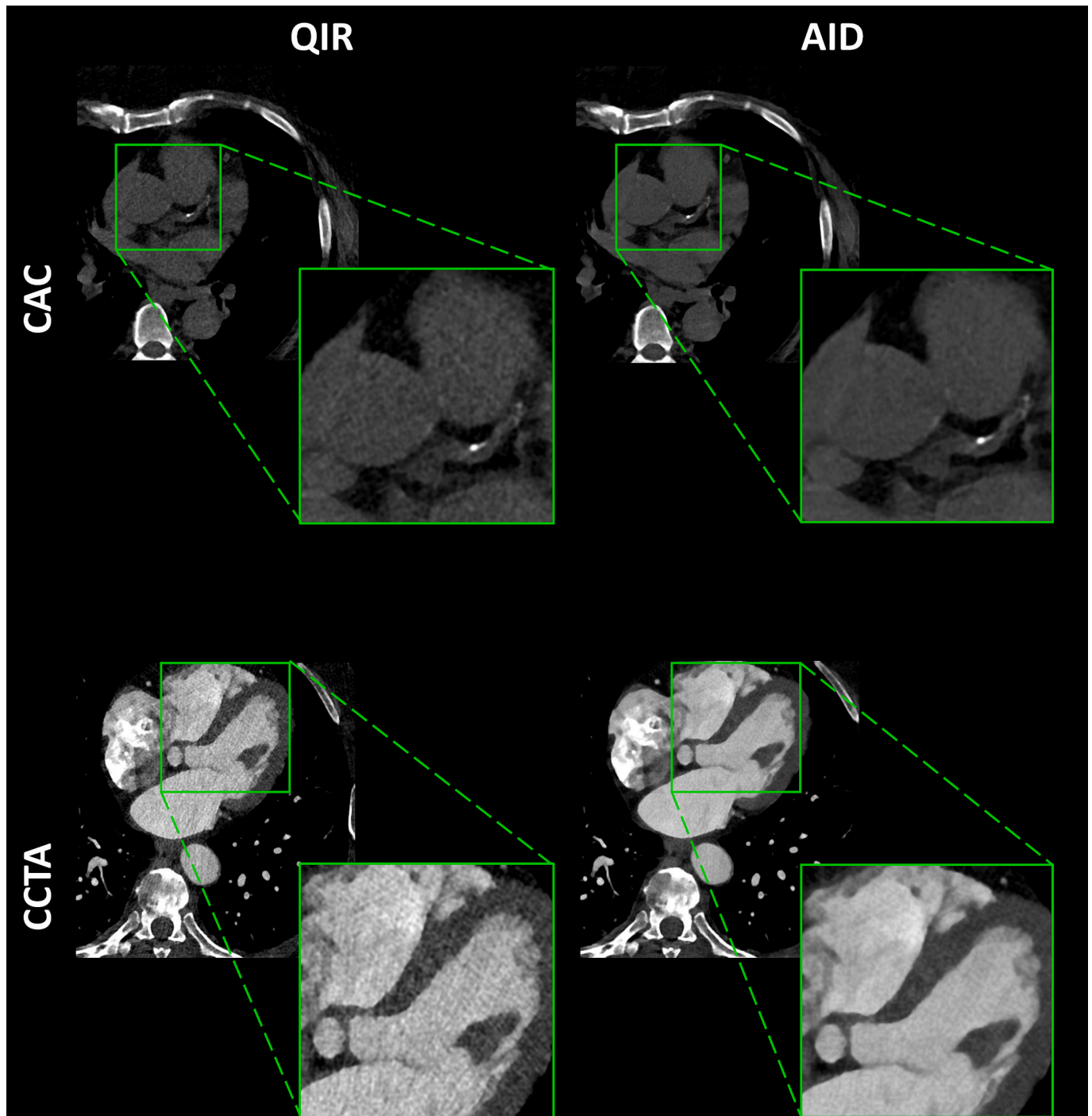


Fig. 5. Representative CAC and CCTA images (QIR left, AID right). Enlarged regions illustrate perceptible differences in overall image clarity, consistent with higher subjective image quality scores for AID.

Table 2

Generalized estimating equations (GEE) model results for subjective image quality.

%	%	%	Coefficient (SE)	%	FDR-adjusted p-value
Subjective Image Quality	CAC	Intercept (QIR at Mean BMI)	-0.137 (0.010)	[-0.157, -0.118]	<0.001
		Main Effect:	0.276 (0.019)	[0.238, 0.314]	<0.001
		AID (vs QIR)	-0.001 (0.002)	[-0.005, 0.003]	0.781
		Main Effect: BMI (centered)	0.000 (0.004)	[-0.008, 0.008]	0.952
		Interaction: AID × BMI	-0.133 (0.010)	[-0.152, -0.114]	<0.001
	CCTA	Intercept (QIR at Mean BMI)	0.264 (0.019)	[0.227, 0.302]	<0.001
		AID (vs QIR)	-0.010 (0.002)	[-0.014, -0.006]	<0.001
		Main Effect: BMI (centered)	0.020 (0.004)	[0.011, 0.028]	<0.001
		Interaction: AID × BMI			

Table 2: Generalized Estimating Equations (GEE) assessing subjective image quality. AID: AI Denoising; BMI: Body Mass Index; CAC: Coronary Artery Calcium; CCTA: Coronary CT Angiography; CI: Confidence Interval; QIR: Quantum Iterative Reconstruction; SE: Robust Standard Error. The Intercept represents the estimated mean for QIR at the cohort's average BMI. The Main Effect of AID represents the mean difference (AID minus QIR) at average BMI. The Interaction term assesses whether the effect of the reconstruction algorithm changes as BMI increases. P-values are adjusted for multiple comparisons using the False Discovery Rate (FDR).

AID reduced the area under the noise power spectrum curve (AUC_NPS) by 164.56 units relative to QIR ($\beta = -164.56$, SE 1.20, 95 % CI: -166.91 to -162.21; $p < 0.001$), corresponding to an approximate 64 % reduction in noise magnitude at the reference BMI. The BMI-dependent increase in noise under AID was 9.67 units per unit of BMI, compared to 17.25 units under QIR, representing a 44 % attenuation of BMI-related noise escalation. The interaction term between BMI and reconstruction confirmed this significantly reduced BMI sensitivity under AID ($\beta = -7.58$, SE 0.90, 95 % CI: -9.34 to -5.81; $p < 0.001$).

3.3.1.2. CCTA datasets. In the contrast-enhanced CCTA protocol, at mean BMI, AID reduced AUC_NPS by 999.62 units relative to QIR ($\beta = -999.62$, SE 15.50, 95 % CI: -1030.01 to -969.24; $p < 0.001$), corresponding to an approximate 78 % reduction in total noise magnitude. The BMI-dependent noise increase under AID was restricted to 27.80 units per BMI unit, compared to 125.67 units under QIR, again reflecting a 78 % attenuation of BMI-driven noise amplification. The interaction term ($\beta = -97.87$, SE 11.60, 95 % CI: -120.61 to -75.14; $p < 0.001$) further demonstrated this substantial mitigation of BMI-driven noise escalation. The interaction between BMI and reconstruction method for noise magnitude (AUC_NPS) in CAC and CCTA datasets is illustrated in Fig. 6. Detailed numerical results are given in Table 3.

3.3.2. Noise texture

3.3.2.1. CAC datasets. The influence of AID on noise texture (PeakF_NPS) differed between imaging protocols. In the CAC protocol ($n = 100$ paired observations), the posterior mean paired difference (QIR - AID) was 0.083 mm⁻¹, with a 95 % Highest Density Interval (HDI) ranging from 0.071 to 0.094 mm⁻¹. The probability of this difference falling within the predefined ROPE was < 0.0001, indicating a measurable shift toward lower spatial frequencies and a smoother noise texture under AID.

3.3.2.2. CCTA datasets. Conversely, in the CCTA protocol ($n = 100$ paired observations), the posterior mean paired PeakF_NPS difference was minimal at 0.004 mm⁻¹ (95 % HDI: 0.004 to 0.004 mm⁻¹). The posterior probability mass falling within the predefined ± 0.008 mm⁻¹ equivalence bounds was > 0.9999. This establishes practical equivalence in peak frequency, indicating that AID preserved the natural noise appearance in contrast-enhanced images. This behavior is illustrated in Fig. 7, which shows the Bayesian posterior distributions of the paired peak-frequency differences between AID and QIR for both CAC and CCTA datasets.

3.3.3. Spatial resolution

3.3.3.1. CAC datasets. In the non-contrast CAC protocol, AID reduced the Edge Rise Distance (ERD) by 0.350 mm at mean BMI relative to QIR ($\beta = -0.350$, SE 0.076, 95 % CI: -0.499 to -0.201; $p < 0.001$), corresponding to a relative reduction of approximately 1.2 %.

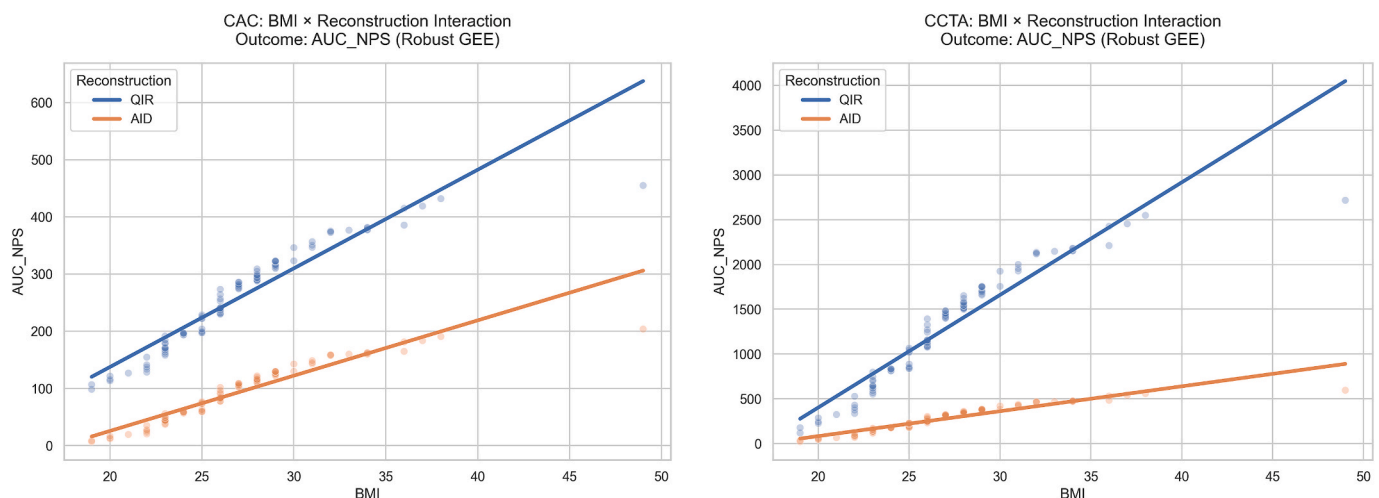


Fig. 6. Modeled effect of BMI on noise magnitude for QIR and AID reconstructions. GEE marginal effect plots for the CAC (left) and CCTA (right). The y-axis represents noise magnitude expressed as the area under the noise power spectrum curve (AUC_NPS), and the x-axis represents patient BMI. Lines indicate the estimated marginal means from robust GEE models. Left: In the non-contrast CAC protocol, noise magnitude increases with BMI under both reconstruction methods, but AID substantially attenuates BMI-related noise escalation compared with QIR. Right: In the contrast-enhanced CCTA protocol, QIR demonstrates a pronounced BMI-dependent increase in noise magnitude, whereas AID markedly reduces both overall noise levels and BMI sensitivity.

Table 3

Generalized estimating equations (GEE) model results for total noise magnitude (AUC_NPS).

Outcome	Dataset	Model term	Coefficient (SE)	95% CI	FDR-adjusted p-value
AUC_NPS	CAC	Intercept (QIR at Mean BMI)	257.298 (2.732)	[251.944, 262.652]	<0.001
		Main Effect: AID (vs QIR)	-164.559 (1.201)	[-166.913, -162.205]	<0.001
		Main Effect: BMI (centered)	17.246 (2.044)	[13.239, 21.252]	<0.001
		Interaction: AID × BMI	-7.576 (0.900)	[-9.340, -5.813]	<0.001
	CCTA	Intercept (QIR at Mean BMI)	1276.582 (19.906)	[1237.567, 1315.596]	<0.001
		Main Effect: AID (vs QIR)	-999.623 (15.503)	[-1030.007, -969.238]	<0.001
		Main Effect: BMI (centered)	125.673 (14.897)	[96.476, 154.870]	<0.001
		Interaction: AID × BMI	-97.874 (11.602)	[-120.613, -75.135]	<0.001

Table 3. Generalized Estimating Equations (GEE) assessing noise magnitude derived from the area under the Noise Power Spectrum curve (AUC_NPS). AID: AI Denoising; AUC: Area Under the Curve; BMI: Body Mass Index; CAC: Coronary Artery Calcium; CCTA: Coronary CT Angiography; CI: Confidence Interval; QIR: Quantum Iterative Reconstruction; SE: Robust Standard Error. The Interaction term assesses whether the noise-reduction effect of AID is modified by increasing BMI. P-values are adjusted for multiple comparisons using the False Discovery Rate (FDR).

Correspondingly, the slope-based resolution surrogate demonstrated a significant increase in edge steepness under AID ($\beta = -0.278$, SE 0.038, 95 % CI: -0.353 to -0.204 ; $p < 0.001$). Regarding body habitus, ERD did not demonstrate a statistically significant interaction between BMI and reconstruction type ($\beta = -0.015$, SE 0.012, 95 % CI: -0.038 to 0.007 ; $p = 0.211$), indicating that the preservation of structural sharpness remained largely stable across patient sizes. However, for slope-based metrics, a significant interaction was observed ($\beta = -0.021$, SE 0.009, 95 % CI: -0.037 to -0.004 ; $p = 0.019$), suggesting a modest

attenuation of BMI-related slope changes under AID.

To determine if these statistically detectable changes in resolution were clinically meaningful, a Bayesian ROPE analysis was performed on the paired observations using a fixed $\pm 2.5\%$ equivalence bound. For ERD in the CAC protocol, the posterior mean relative difference (QIR – AID) was -0.012 (95 % HDI: -0.016 to -0.009). The posterior probability of falling within the $\pm 2.5\%$ ROPE was > 0.9999 . For edge slope, the posterior mean relative increase was 0.011 (95 % HDI: 0.008 to 0.013), with a ROPE probability of > 0.9999 . This indicates that relative changes remained entirely within predefined practical equivalence limits despite statistical detectability.

3.3.3.2. CCTA datasets. In the contrast-enhanced CCTA protocol, AID similarly reduced ERD by 0.058 mm at mean BMI ($\beta = -0.058$, SE 0.021, 95 % CI: -0.100 to -0.016 ; $p = 0.008$), again indicating sharper structural transitions under AID. Edge slope was also significantly steeper under AID compared to QIR ($\beta = -2.130$, SE 0.453, 95 % CI: -3.017 to -1.242 ; $p < 0.001$). Similar to the CAC protocol, ERD did not demonstrate a significant interaction between BMI and reconstruction type ($\beta = 0.016$, SE 0.009, 95 % CI: -0.001 to 0.033 ; $p = 0.077$). Edge slope also showed no significant BMI interaction in the CCTA dataset ($\beta = 0.165$, SE 0.146, 95 % CI: -0.121 to 0.452 ; $p = 0.287$).

The posterior mean relative ERD difference was -0.015 (95 % HDI: -0.022 to -0.007), with a probability of 0.9945 of falling within the ROPE. For edge slope, the posterior mean relative increase was 0.017 (95 % HDI: 0.012 to 0.023), with a ROPE probability of 0.9940 . Overall, the pooled posterior mean relative difference across both protocols ($n = 200$) was -0.014 for ERD and 0.013 for edge slope, with overall ROPE probabilities of > 0.9999 . These findings confirm that although structural sharpness is statistically improved under AID, the magnitude of alteration is practically equivalent to QIR, ensuring no structural boundaries are artificially smoothed or generated. The relationship between BMI, reconstruction method, and spatial resolution metrics is illustrated in Fig. 8. The data regarding GEE Model results for spatial resolution can be found in Table 4. Fig. 9 visually demonstrates reduced noise and enhanced sharpness with AID.

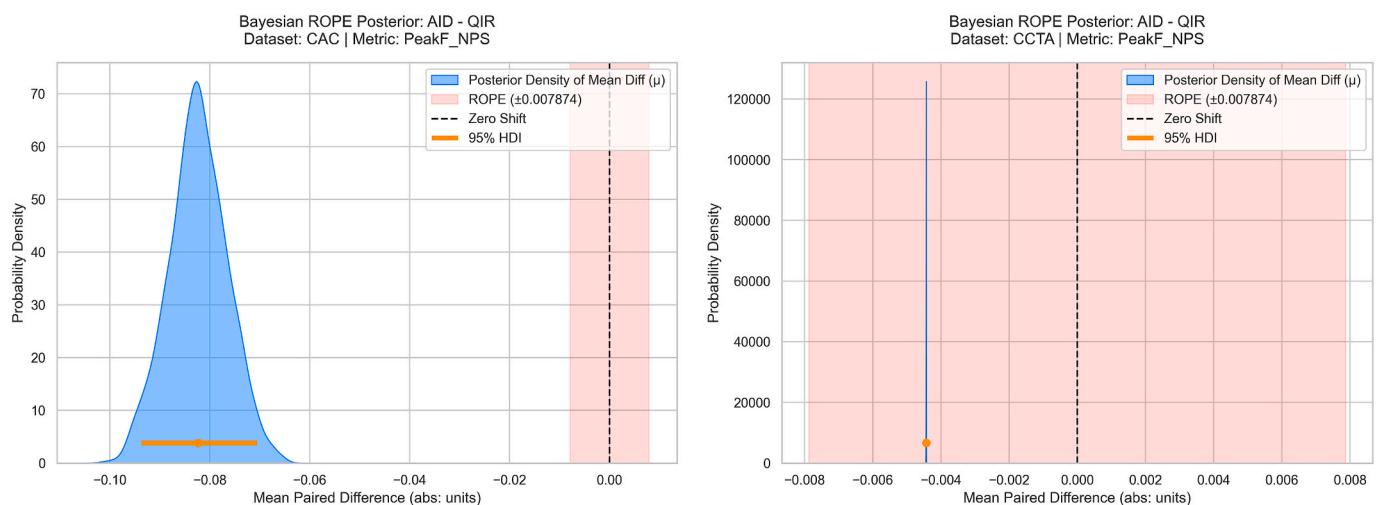


Fig. 7. Modeled effect of reconstruction method on noise texture expressed as peak spatial frequency (PeakF_NPS). Bayesian posterior distributions of the mean paired difference between AID and QIR reconstructions for CAC (left) and CCTA (right). The x-axis represents the mean paired difference in peak spatial frequency (mm^{-1}), and the y-axis represents the posterior probability density. Blue curves indicate the posterior distribution, orange bars represent the 95 % HDI, and the red shaded area denotes the predefined ROPE. Left: In the CAC protocol, the posterior distribution lies clearly outside the ROPE, indicating a measurable shift toward lower spatial frequencies under AID and a smoother noise texture. Right: In the CCTA protocol, the posterior distribution falls entirely within the ROPE, demonstrating practical equivalence of noise texture between AID and QIR reconstructions.

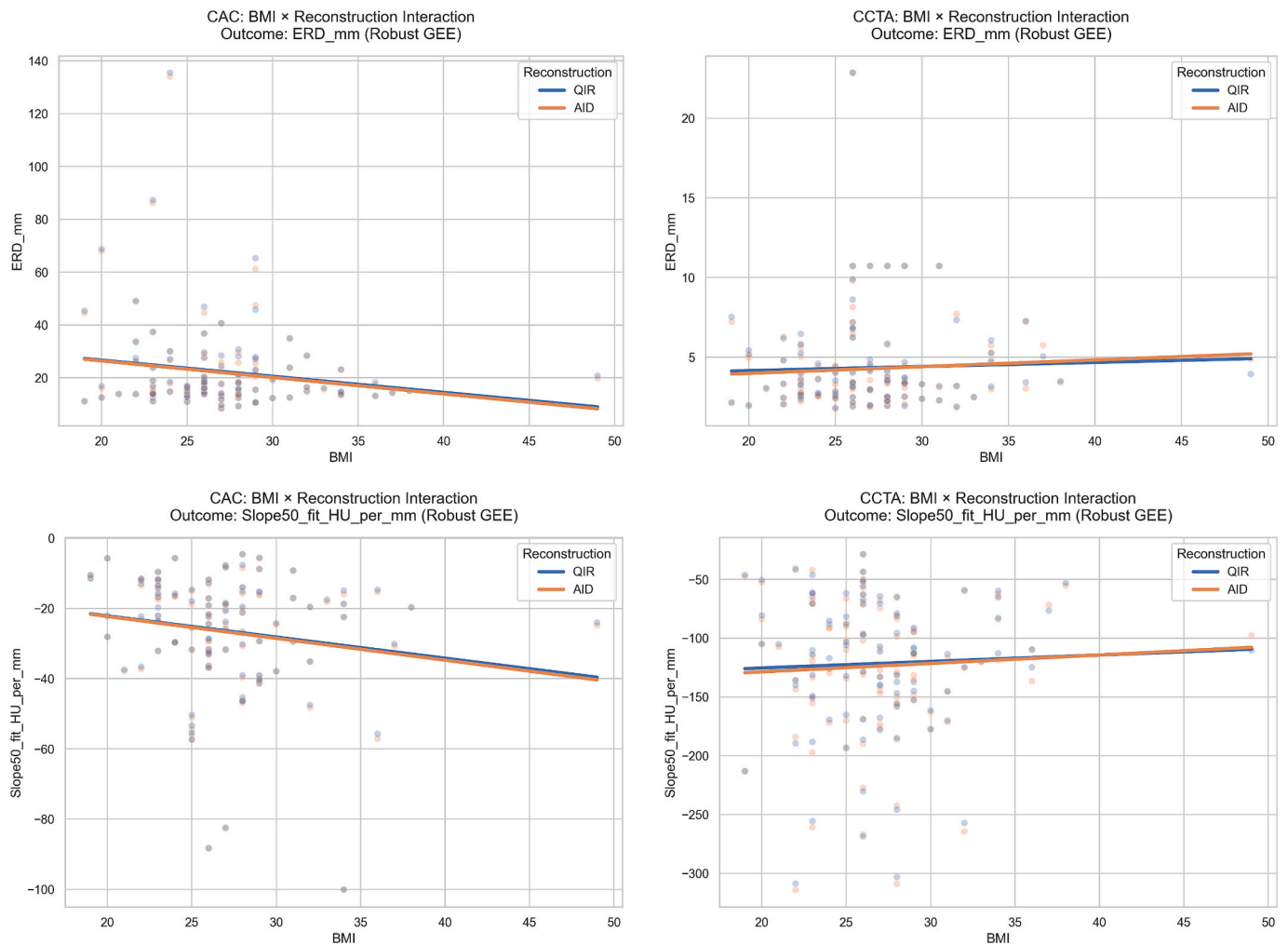


Fig. 8. Modeled effect of BMI on spatial resolution metrics for QIR and AID reconstructions. GEE marginal effect plots for the CAC (left column) and CCTA (right column). The x-axis represents patient BMI. Top row: ERD (mm) as a measure of edge transition width. Bottom row: Slope (edge steepness, HU/mm) as a surrogate of spatial resolution. Lines indicate estimated marginal means from robust GEE models and points represent individual observations. Left: In the CAC protocol, AID demonstrates slightly reduced ERD and steeper edge slopes compared with QIR, indicating modestly sharper edge transitions without a significant BMI-dependent interaction for ERD. Right: In contrast-enhanced CCTA protocol, AID similarly shows slightly lower ERD and steeper slopes relative to QIR, while spatial resolution metrics remain largely stable across patient BMI.

3.4. Clinical interchangeability

3.4.1. CAC datasets

For continuous Agatston scores prior to manual correction, QIR and AID yielded highly comparable mean values (256.6 ± 696.3 vs. 262.7 ± 698.2 , respectively). The posterior mean relative difference was minimal at -0.003 (95 % HDI: -0.006 to 0.000), with the probability of practical equivalence exceeding 0.999 . Following manual correction, Agatston scores remained practically equivalent (QIR: 266.0 ± 715.5 ; AID: 270.7 ± 714.6 ; mean difference: -0.004), maintaining a ROPE probability of > 0.999 . This demonstrates that AID preserves standard calcium quantification without introducing systematic scaling bias.

For the categorical assignment of coronary age, diagnostic equivalence was stringently defined as a latent shift of less than one full clinical age category. Prior to manual correction, the overall posterior mean difference was 0.815 (95 % HDI: 0.280 to 1.446), yielding a 0.744 probability of practical equivalence. Following routine manual correction, alignment between the reconstructions improved, reducing the latent mean difference to 0.657 (95 % HDI: 0.185 to 1.188) and increasing the probability of overall practical equivalence to 0.916 . Notably, within specific individual coronary age brackets, the probability of practical equivalence remained exceptionally high both before

and after correction (ranging from 0.910 to > 0.999). These findings indicate that while minor sub-category variations occur before manual correction, AID does not cause clinically meaningful shifts in patient risk categorization. The numerical values are provided in [Table 5](#).

3.4.2. CCTA datasets

For CAD-RADS scoring, the overall posterior mean latent difference between QIR and AID was minimal at -0.002 (95 % HDI: -0.169 to 0.167), yielding a probability of practical equivalence of > 0.999 . This exceptionally high probability of practical equivalence was consistently observed across all primary coronary branches (RCA, LM, LAD, and LCx) and across all individual CAD-RADS severity categories (0 through 4/5). This confirms that the robust noise reduction achieved by AID does not systematically shift patients across CAD-RADS risk categories, thereby preserving established clinical decision-making pathways. Detailed numerical results are given in [Table 6](#).

Similarly, for plaque characterization, diagnostic equivalence remained high. The overall latent mean difference was 0.001 (95 % HDI: -0.035 to 0.037), with a probability of practical equivalence of > 0.999 . Analysis of individual plaque composition categories (non-calcified, mixed, and calcified plaques) across the overall system and within individual coronary branches also demonstrated near-perfect practical

Table 4
Generalized Estimating Equations (GEE) model results for spatial resolution.

Outcome	Dataset	Model TERM	Coefficient (SE)	95% CI	FDR-adjusted p-value
Edge Rise Distance (mm)	CAC	Intercept (QIR at Mean BMI)	22.426 (1.686)	[19.122, 25.731]	<0.001
		Main Effect: AID (vs QIR)	−0.350 (0.076)	[−0.499, −0.201]	<0.001
		Main Effect: BMI (centered)	−0.610 (0.333)	[−1.263, 0.042]	0.081
		Interaction: AID × BMI	−0.015 (0.012)	[−0.038, 0.007]	0.211
	CCTA	Intercept (QIR at Mean BMI)	4.339 (0.288)	[3.775, 4.903]	<0.001
		Main Effect: AID (vs QIR)	−0.058 (0.021)	[−0.100, −0.016]	0.008
		Main Effect: BMI (centered)	0.026 (0.038)	[−0.049, 0.101]	0.536
		Interaction: AID × BMI	0.016 (0.009)	[−0.001, 0.033]	0.077
Edge Slope (HU/mm)	CAC	Intercept (QIR at Mean BMI)	−26.335 (1.653)	[−29.574, −23.095]	<0.001
		Main Effect: AID (vs QIR)	−0.278 (0.038)	[−0.353, −0.204]	<0.001
		Main Effect: BMI (centered)	−0.602 (0.385)	[−1.356, 0.152]	0.138
		Interaction: AID × BMI	−0.021 (0.009)	[−0.037, −0.004]	0.019
	CCTA	Intercept (QIR at Mean BMI)	−121.533 (5.775)	[−132.851, −110.215]	<0.001
		Main Effect: AID (vs QIR)	−2.130 (0.453)	[−3.017, −1.242]	<0.001
		Main Effect: BMI (centered)	0.552 (1.116)	[−1.634, 2.738]	0.653
		Interaction: AID × BMI	0.165 (0.146)	[−0.121, 0.452]	0.287

Table 4. Generalized Estimating Equations (GEE) assessing spatial resolution. AID: AI Denoising; BMI: Body Mass Index; CAC: Coronary Artery Calcium; CCTA: Coronary CT Angiography; CI: Confidence Interval; QIR: Quantum Iterative Reconstruction; SE: Robust Standard Error. The Intercept represents the estimated mean for QIR at the cohort's average BMI. The Main Effect of AID represents the mean difference (AID minus QIR) at average BMI. P-values are adjusted for multiple comparisons using the False Discovery Rate (FDR).

equivalence, with probabilities ranging from 0.999 to > 0.999. These findings establish that AID reliably maintains the critical structural and density-based features required for accurate coronary plaque assessment and downstream patient management. The underlying numerical data are presented in detail in Table 7.

4. Discussion

This study evaluated whether a PCD-CT weighted AI-based denoising algorithm improves subjective and objective image quality regarding residual BMI-related noise in CAC and CCTA while preserving quantitative accuracy and workflow efficiency compared with standard QIR reconstruction. Our results demonstrate consistent improvements in visual image quality and objective metrics, including ERD and NPS, without affecting attenuation values, coronary risk scores, lesion characteristics, or manual correction times. These findings suggest that AID can be considered a neutral yet beneficial enhancement: it improves visual and measurable image quality without biasing clinical readouts or workflow, supporting its clinical interchangeability in cardiac PCD-CT.

Our findings demonstrated significant improvements in subjective image quality for both CAC and CCTA when using AID. Radiologists consistently reported enhanced overall clarity with AID datasets. This is consistent with prior research showing that deep learning-based denoising improves visual quality and diagnostic confidence in cardiac CT, as described by Benz et al. [10]. Similarly, Wang et al. [11], Mei et al. [12], and Tatsugami et al. [13] observed improved subjective image quality in CCTA using deep learning reconstruction. However, these studies analyzed only CCTA datasets and did not include CAC. Moreover, none employed PCD-CT algorithms, limiting direct comparability. The forced-choice paired-comparison design demonstrated strong and consistent reader preference for AID across both CAC and CCTA protocols. The magnitude of the reconstruction effect was large relative to the bounded preference scale and remained significant after accounting for repeated measures within patients. Importantly, subjective findings paralleled objective metrics. In CCTA datasets, BMI-increases were associated with reduced subjective quality under QIR, consistent with objective noise escalation observed in NPS analysis. AID significantly mitigated this BMI-related penalty both objectively and subjectively. In CAC datasets, subjective preference for AID was strong but largely BMI-independent, aligning with objective findings of robust noise suppression without practical degradation of spatial resolution. Taken together with the objective analyses of noise magnitude, noise

texture, and edge sharpness, the subjective results support the conclusion that AID improves perceived image quality while maintaining structural delineability and exhibiting robustness across patient habitus, in line with prior reports demonstrating improved image quality across BMI categories with deep learning reconstruction [14]. In summary, our findings expand existing evidence, supporting deep learning's potential to improve subjective image quality across cardiac CT applications.

In objective quality assessments, AID significantly improved Noise magnitude in both CAC and CCTA. Prior studies, including Mei et al. [12] and Hong et al. [15] reported similar benefits of deep learning reconstruction in conventional CCTA, but all were limited to energy-integrating detectors. Nagayama et al. [16] and Funama et al. [17] emphasized NPS analysis for noise quantification, though their work was restricted to phantoms or non-cardiac CT scans. Recently, a cardiac PCD-CT study [18] combining iterative reconstruction with a convolutional neural network-based post-hoc denoising reported improved signal-to-noise ratio and contrast-to-noise ratio while preserving diagnostic performance in myocardial computed tomography late enhancement. However, that work focused primarily on myocardial tissue characterization rather than coronary CT applications and used high-level noise metrics instead of in-depth characterizations. Importantly, high BMI was associated with increased residual noise under conventional reconstruction, reflecting the well-established impact of photon starvation on image quality. AID markedly reduced BMI-dependent noise escalation, with relative reductions of approximately 44% in CAC and 78% in CCTA. These effects indicate improved robustness of image quality across varying patient habitus. Although this study was not designed to formally evaluate dose reduction strategies, the reduced BMI sensitivity under AID suggests potential for more stable image quality at lower tube current settings, particularly in larger patients.

The influence of AID on noise texture was protocol dependent. In CAC acquisitions, AID induced a shift toward lower spatial frequencies consistent with smoothing behavior. In contrast, CCTA acquisitions demonstrated practical equivalence in peak frequency, with any shift remaining below half a frequency bin. This differential behavior suggests that texture preservation under deep learning reconstruction is context-specific and may depend on contrast conditions, dose level, or reconstruction kernel characteristics during training phase rather than representing a uniform alteration of noise properties. Importantly, dedicated investigations of AI-based denoising specifically targeting texture in coronary PCD-CT remain limited.

Regarding spatial resolution analysis, prior studies have placed

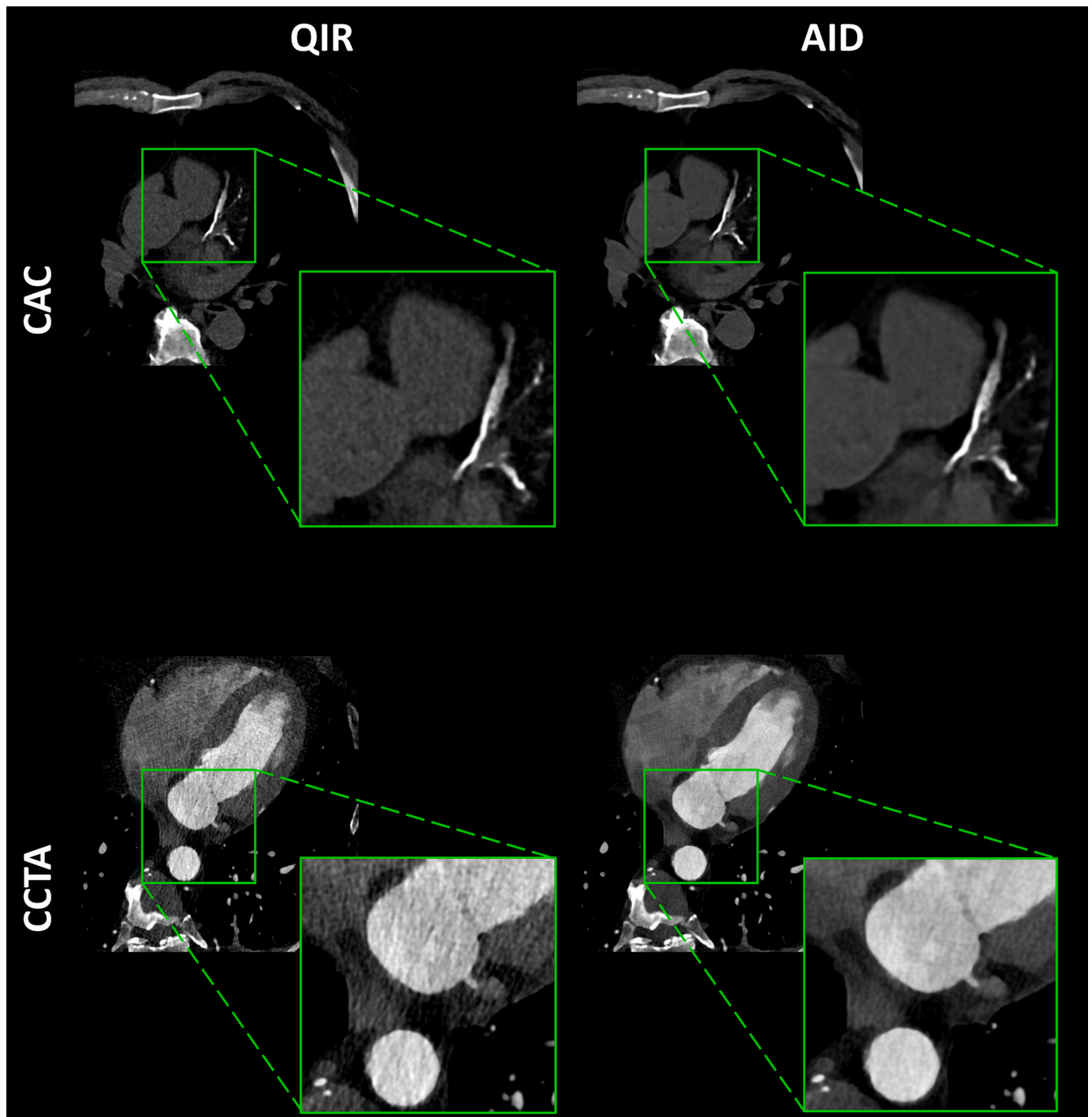


Fig. 9. Representative CAC (top row) and CCTA (bottom row) images reconstructed with QIR (left column) and AID (right column). Enlarged regions of interest demonstrate reduced image noise and improved vessel delineation with AID compared to QIR.

objective ROIs directly within coronary segments [19], whereas we intentionally positioned our ERD ROIs in stable surrogate tissues. This strategy was selected to ensure reproducible measurements in tissue with attenuation characteristics comparable to coronary targets while reducing susceptibility to motion artifacts and partial volume effects that commonly affect small coronary ROIs. Edge spread analysis indicates that AID does not compromise spatial resolution and may yield detectable small, but practically negligible improvements in edge sharpness. In contrast to historical reconstruction trade-offs, where aggressive noise suppression may introduce smoothing effects and alter spatial resolution characteristics as previously described for iterative

reconstruction techniques [20], the present findings demonstrate that AID reduces noise burden without inducing degradation of spatial resolution. These effects also remained stable across varying patient body habitus, reinforcing the robustness of the reconstruction approach. Furthermore, the non-linear curve fitting utilized for edge response analysis intrinsically modeled the absolute physical attenuation boundaries. The baseline tissue attenuation (minimum HU) and the contrast step height (maximum HU difference) were directly parameterized within the logistic function. Because the analytical derivation of the edge slope is strictly dependent on this step height, the stability and practical equivalence of these spatial metrics concurrently confirm that

Table 5
Bayesian diagnostic equivalence analysis for coronary artery calcium (CAC) scoring metrics.

Metric	Correction status	Category	QIR	AID	Mean Diff. [95% HDI]	Prob. in ROPE
Agatston Score	<i>Before Correction</i>	<i>Continuous</i>	256.6 ± 696.3	262.7 ± 698.2	−0.003 [−0.006, −0.000]	>0.999
	<i>After Correction</i>	<i>Continuous</i>	266.0 ± 715.5	270.7 ± 714.6	−0.004 [−0.007, −0.001]	>0.999
Coronary Age	<i>Before Correction</i>	Overall	—	—	0.815 [0.280, 1.446]	0.744
		<35 years	46 (46.0%)	41 (41.0%)	−0.610 [−1.032, −0.211]	0.965
		40–44 years	4 (4.0%)	6 (6.0%)	−0.048 [−0.462, 0.349]	>0.999
		45–49 years	4 (4.0%)	2 (2.0%)	−0.473 [−0.904, −0.059]	0.992
		50–54 years	6 (6.0%)	6 (6.0%)	−0.268 [−0.687, 0.131]	>0.999
		55–59 years	7 (7.0%)	11 (11.0%)	0.149 [−0.240, 0.512]	>0.999
		60–64 years	5 (5.0%)	4 (4.0%)	−0.354 [−0.768, 0.050]	0.999
		>70 years	28 (28.0%)	30 (30.0%)	0.106 [−0.279, 0.505]	>0.999
	<i>After Correction</i>	Overall	—	—	0.657 [0.185, 1.188]	0.916
		<35 years	44 (44.0%)	38 (38.0%)	−0.707 [−1.127, −0.313]	0.910
		40–44 years	4 (4.0%)	8 (8.0%)	0.135 [−0.258, 0.520]	>0.999
		45–49 years	5 (5.0%)	4 (4.0%)	−0.322 [−0.735, 0.067]	>0.999
		50–54 years	6 (6.0%)	6 (6.0%)	−0.268 [−0.687, 0.131]	>0.999
		55–59 years	8 (8.0%)	10 (10.0%)	−0.030 [−0.411, 0.342]	>0.999
		60–64 years	5 (5.0%)	4 (4.0%)	−0.354 [−0.768, 0.050]	0.999
		>70 years	28 (28.0%)	30 (30.0%)	0.106 [−0.279, 0.505]	>0.999

Table 5: Analysis of continuous Agatston Scores and categorical Coronary Age assignments derived from calcium scoring scans, evaluated before and after specific image domain corrections. Abbreviations: AID: AI Denoising; HDI: Highest Density Interval; QIR: Quantum Iterative Reconstruction; ROPE: Region of Practical Equivalence. Data presentation: For the continuous Agatston Score metric, values under QIR and AID are presented as Mean Standard Deviation. For the categorical Coronary Age metric, values are presented as counts with percentages (%). Statistical parameters: Mean Diff.: The posterior mean of the differences between the paired measurements. 95 % HDI: The 95 % Highest Density Interval, representing the range within which the true difference falls with 95 % probability. ROPE Bounds: The predefined practical equivalence thresholds. Prob. in ROPE: The posterior probability that the true mean difference lies entirely within the practical equivalence bounds of ± 2.5 % for continuous variables and below one latent category shift for categorical variables. ROPE Values are formatted as proportions rather than percentages, consistent with standard statistical reporting, where >0.999 denotes near-certainty.

AID preserved the underlying CT density values without introducing artificial density shifts. As highlighted by Chandran et al. [21] and Delabie et al. [22], attenuation comparability is a fundamental prerequisite when validating reconstruction algorithms. In our cohort, no significant quantitative deviations were observed between QIR and AID for CAC scoring, indicating that advanced denoising does not bias established clinical risk measures [23]. Prior studies in EID-CT have demonstrated that AID algorithms can reduce image noise without altering mean CT attenuation [24,25]. However, traditional studies often relied on simple non-inferiority testing, and none addressed the unique spectral properties of PCD-CT. By utilizing a rigorous Bayesian ROPE framework, our results provide mathematical evidence that a PCD-CT weighted AID algorithm perfectly preserves the strict >130 HU threshold required for Agatston scoring. Furthermore, our CCTA analysis focused strictly on relevant diagnostic outcomes rather than secondary operational metrics. While prior literature validating conventional AID algorithms has heavily emphasized reader workflow efficiency and manual correction times [26–28], demonstrating that sharper images do not increase reader burden is only a secondary benefit. The primary prerequisite for clinical adoption is diagnostic interchangeability. Our Bayesian analysis proved a near-certain probability of practical equivalence across all CAD-RADS severity categories and coronary branches. This confirms that the robust noise reduction achieved by AID does not systematically shift patients across CAD-RADS risk categories, thereby safely preserving established clinical decision-making pathways. Finally, we extended our evaluation to the exact density-based structural features that define modern cardiac CT. Earlier validations of AID algorithms often relied on macroscopic functional parameters, such as left ventricular volume or myocardial mass, to establish quantitative safety [29]. While useful, the true diagnostic challenge in PCD-CT is the microscopic evaluation of coronary plaque. By demonstrating near-certain diagnostic equivalence for specific plaque composition categories, our study provides the first evidence that PCD-CT AID preserves both the geometric boundaries and the absolute intraluminal attenuation profiles necessary for advanced plaque characterization. In summary, this PCD-CT AID algorithm improves subjective and objective image quality while maintaining stringent diagnostic

equivalence, confirming clinical interchangeability with QIR across the diagnostic spectrum of cardiac CT.

Our study has several limitations that warrant consideration. First, the retrospective, single-center design restricts generalizability, and inherent selection bias cannot be fully excluded. Prospective, multi-center validation is necessary to confirm our findings across diverse patient populations.

Second, the absence of an invasive reference standard, such as invasive coronary angiography (ICA) or optical coherence tomography (OCT), precludes the definitive assessment of absolute diagnostic accuracy. Our Bayesian equivalence framework successfully proves strict clinical interchangeability with QIR, but future studies utilizing invasive ground truths are required to determine if AID improves actual diagnostic performance.

Third, although our clinical interchangeability endpoints were highly comprehensive, spanning Agatston scoring, coronary age, CAD-RADS, and plaque characterization, they do not encompass the entirety of cardiac CT applications. Future investigations should incorporate functional parameters, such as CT-derived fractional flow reserve (CT-FFR) or advanced radiomic profiling, to assess the broader downstream impact of AI denoising.

Fourth, parameters surrounding the image acquisition and baseline reconstruction pose inherent limits. Image reconstruction was performed at a spatial resolution of 0.4 mm based on standard retrospective clinical protocols. Because ultra-high-resolution (UHR) settings (e.g., 0.2 mm) are a hallmark of PCD-CT and strongly influence stenosis assessment, algorithm performance under UHR conditions requires targeted reassessment. Furthermore, QIR was applied at a standardized, non-maximal strength level prior to AID application. While reflective of routine clinical practice, evaluating AID against various baseline QIR strengths is needed to identify optimal combinatorial reconstruction strategies.

Fifth, while the AID algorithm is designed to be vendor-neutral, our investigation was performed exclusively within a single-vendor PCD-CT environment. As PCD-CT technology expands, generalizability across different hardware ecosystems must be verified. Similarly, objective physics evaluations (NPS and ESF/ERD) were performed using

Table 6
Bayesian diagnostic equivalence analysis for CAD-RADS classifications.

Coronary Artery	Category	QIR	AID	Mean Diff. [95% HDI]	Prob. in ROPE
Overall System	Overall	—	—	−0.002 [−0.169, 0.167]	>0.999
	CAD-RADS 0	1407 (82.8%)	1407 (82.8%)	0.018 [−0.109, 0.143]	>0.999
	CAD-RADS 1	158 (9.3%)	158 (9.3%)	−0.031 [−0.168, 0.103]	>0.999
	CAD-RADS 2	48 (2.8%)	48 (2.8%)	−0.121 [−0.328, 0.085]	>0.999
	CAD-RADS 3	50 (2.9%)	50 (2.9%)	−0.135 [−0.336, 0.056]	>0.999
	CAD-RADS 4/5	36 (2.1%)	36 (2.1%)	−0.155 [−0.367, 0.049]	>0.999
RCA	Overall	—	—	−0.013 [−0.247, 0.218]	>0.999
	CAD-RADS 0	413 (82.6%)	413 (82.6%)	0.040 [−0.200, 0.279]	>0.999
	CAD-RADS 1	51 (10.2%)	51 (10.2%)	−0.067 [−0.317, 0.168]	>0.999
	CAD-RADS 2	14 (2.8%)	14 (2.8%)	−0.239 [−0.571, 0.087]	>0.999
	CAD-RADS 3	14 (2.8%)	14 (2.8%)	−0.262 [−0.597, 0.068]	>0.999
	CAD-RADS 4/5	8 (1.6%)	8 (1.6%)	−0.367 [−0.742, −0.008]	>0.999
LM	Overall	—	—	−0.060 [−0.414, 0.284]	>0.999
	CAD-RADS 0	170 (85.0%)	170 (85.0%)	0.095 [−0.254, 0.439]	>0.999
	CAD-RADS 1	22 (11.0%)	22 (11.0%)	−0.175 [−0.592, 0.222]	>0.999
	CAD-RADS 2	5 (2.5%)	5 (2.5%)	−0.330 [−0.723, 0.024]	>0.999
	CAD-RADS 3	3 (1.5%)	3 (1.5%)	−0.409 [−0.801, −0.046]	0.998
LAD	Overall	—	—	−0.002 [−0.252, 0.263]	>0.999
	CAD-RADS 0	378 (75.8%)	378 (75.8%)	0.023 [−0.208, 0.245]	>0.999
	CAD-RADS 1	56 (11.2%)	56 (11.2%)	−0.054 [−0.298, 0.194]	>0.999
	CAD-RADS 2	23 (4.6%)	23 (4.6%)	−0.158 [−0.458, 0.142]	>0.999
	CAD-RADS 3	22 (4.4%)	22 (4.4%)	−0.154 [−0.435, 0.110]	>0.999
	CAD-RADS 4/5	20 (4.0%)	20 (4.0%)	−0.193 [−0.474, 0.081]	>0.999

Table 6 (continued)

Coronary Artery	Category	QIR	AID	Mean Diff. [95% HDI]	Prob. in ROPE
LCx	Overall	—	—	−0.043 [−0.440, 0.332]	>0.999
	CAD-RADS 0	446 (89.2%)	446 (89.2%)	0.055 [−0.181, 0.289]	>0.999
	CAD-RADS 1	29 (5.8%)	29 (5.8%)	−0.107 [−0.366, 0.149]	>0.999
	CAD-RADS 2	6 (1.2%)	6 (1.2%)	−0.390 [−0.745, −0.057]	0.999
	CAD-RADS 3	11 (2.2%)	11 (2.2%)	−0.277 [−0.604, 0.045]	>0.999
	CAD-RADS 4/5	8 (1.6%)	8 (1.6%)	−0.331 [−0.659, −0.018]	>0.999

Table 6: Analysis of categorical CAD-RADS assignments evaluated across primary coronary branches. Abbreviations: AID: AI Denoising; CAD-RADS: Coronary Artery Disease Reporting and Data System; HDI: Highest Density Interval; LAD: Left Anterior Descending; LCx: Left Circumflex; LM: Left Main; QIR: Quantum Iterative Reconstruction; RCA: Right Coronary Artery; ROPE: Region of Practical Equivalence. Data presentation: Values under QIR and AID are presented as counts with percentages (%). Statistical parameters: Mean Diff.: The posterior mean of the differences between the paired measurements on the latent scale. 95 % HDI: The 95 % Highest Density Interval, representing the range within which the true difference falls with 95 % probability. Prob. in ROPE: The posterior probability that the true mean difference lies entirely within the practical equivalence bounds (defined as below one latent category shift). Values are formatted as proportions rather than percentages, consistent with standard statistical reporting, where >0.999 denotes near-certainty.

established ROI-based sampling in selected anatomical regions. Expanding this to automated, whole-volume structural assessments may yield a more nuanced understanding of regional denoising behavior.

Finally, while we mitigated subjective rater bias by employing four independent readers and robust GEE modeling, subjective image quality assessments always carry a degree of residual perception bias. Additionally, AID processing was performed in an offline research environment. Although average processing times were short (approximately 3 to 5 min per dataset), seamless integration into real-time, high-throughput clinical workflows remains to be validated prior to routine deployment.

In conclusion, the PCD-CT-weighted AID algorithm significantly improves subjective and objective image quality in both CAC and CCTA. Most notably, AID effectively attenuates the severe BMI-related noise degradation that traditionally limits conventional reconstructions in larger patients. By utilizing a rigorous Bayesian framework, we demonstrate that these substantial noise reductions are achieved while strictly maintaining quantitative diagnostic equivalence and absolute clinical interchangeability across Agatston scoring, coronary age, CAD-RADS classification, and coronary plaque characterization.

CRediT authorship contribution statement

Lisa Urban: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Josephine Berger:** Writing – review & editing, Validation, Data curation. **Jan M. Brendel:** Writing – review & editing, Validation. **Judith Herrmann:** Writing – review & editing, Validation, Data curation. **Patrick Krumm:** Writing – review & editing, Validation, Resources, Data curation. **Konstantin Nikolaou:** Writing – review & editing, Resources. **Saif Afat:** Writing – review & editing, Validation, Resources, Project administration, Conceptualization. **Andreas Brendlin:** Writing – review & editing, Visualization,

Table 7
Bayesian diagnostic equivalence analysis for coronary plaque characterization.

Coronary Artery	Category	QIR	AID	Mean Diff. [95% HDI]	Prob. in ROPE
Overall System	Overall	—	—	0.001 [−0.035, 0.037]	>0.999
	Non-calcified	47 (16.2%)	47 (16.2%)	−0.059 [−0.331, 0.202]	>0.999
	Mixed	110 (37.8%)	110 (37.8%)	−0.020 [−0.245, 0.205]	>0.999
	Calcified	134 (46.0%)	134 (46.0%)	−0.003 [−0.226, 0.211]	>0.999
RCA	Overall	—	—	0.011 [−0.088, 0.112]	>0.999
	Non-calcified	21 (24.1%)	21 (24.1%)	−0.078 [−0.480, 0.316]	>0.999
	Mixed	20 (23.0%)	20 (23.0%)	−0.089 [−0.451, 0.269]	>0.999
	Calcified	46 (52.9%)	46 (52.9%)	0.012 [−0.351, 0.375]	>0.999
LM	Overall	—	—	−0.000 [−0.081, 0.080]	>0.999
	Non-calcified	2 (6.7%)	2 (6.7%)	−0.175 [−0.606, 0.253]	0.999
	Mixed	16 (53.3%)	16 (53.3%)	0.003 [−0.425, 0.419]	>0.999
	Calcified	12 (40.0%)	12 (40.0%)	−0.021 [−0.454, 0.402]	>0.999
LAD	Overall	—	—	0.003 [−0.043, 0.050]	>0.999
	Non-calcified	16 (13.3%)	16 (13.3%)	−0.164 [−0.557, 0.214]	>0.999
	Mixed	54 (45.0%)	54 (45.0%)	−0.034 [−0.385, 0.312]	>0.999
	Calcified	50 (41.7%)	50 (41.7%)	0.014 [−0.352, 0.363]	>0.999
LCx	Overall	—	—	0.002 [−0.050, 0.056]	>0.999
	Non-calcified	8 (14.8%)	8 (14.8%)	−0.149 [−0.561, 0.270]	>0.999
	Mixed	20 (37.0%)	20 (37.0%)	−0.055 [−0.476, 0.370]	>0.999
	Calcified	26 (48.1%)	26 (48.1%)	0.003 [−0.411, 0.402]	>0.999

Table 7: Analysis of categorical plaque characterization evaluated across primary coronary branches. Abbreviations: AID: AI Denoising; HDI: Highest Density Interval; LAD: Left Anterior Descending; LCx: Left Circumflex; LM: Left Main; QIR: Quantum Iterative Reconstruction; RCA: Right Coronary Artery; ROPE: Region of Practical Equivalence. Data presentation: Values under QIR and AID are presented as counts with percentages (%). Statistical parameters: Mean Diff.: The posterior mean of the differences between the paired measurements on

the latent scale. 95 % HDI: The 95 % Highest Density Interval, representing the range within which the true difference falls with 95 % probability. Prob. in ROPE: The posterior probability that the true mean difference lies entirely within the practical equivalence bounds (defined as below one latent category shift). Values are formatted as proportions rather than percentages, consistent with standard statistical reporting, where >0.999 denotes near-certainty.

Validation, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejrad.2026.112921>.

References

- [1] S.S. Virani, A. Alonso, E.J. Benjamin, et al., Heart disease and stroke statistics-2020 UPDATE: A REPORT FROM THE AMERICAN HEART ASSOCIATION, *Circulation* 141 (9) (2020) e139–e596, <https://doi.org/10.1161/CIR.0000000000000757>.
- [2] G. Herold, *Innere Medizin 2025*, De Gruyter, 2025.
- [3] W. Piper, *Innere Medizin*, Springer, Berlin Heidelberg, 2013.
- [4] J.M. Brendel, K. Nikolaou, B. Foldyna, Plaque-Charakterisierung und individualisierte Risikoeinschätzung, *Radiologie (Heidelb)*. 64 (12) (2024) 946–955, <https://doi.org/10.1007/s00117-024-01385-y>.
- [5] V. Zylka-Menhorn, Europäischer Kardiologenkongress: die koronare Herzerkrankung ist „alles andere als stabil“, *Deutsches Ärzteblatt*. 116 (38) (2019) 8344.
- [6] A. Sharma, M.G. Cerdas, S. Reza-Soltani, et al., A review of photon-counting computed tomography (PCCT) in the diagnosis of cardiovascular diseases, *Cureus* 16 (11) (2024) e73119, <https://doi.org/10.7759/cureus.73119>.
- [7] R. Obuchowicz, M. Strzelecki, A. Piórkowski, Clinical applications of artificial intelligence in medical imaging and image processing-a review, *Cancers (Basel)* 16 (10) (2024), <https://doi.org/10.3390/cancers16101870>.
- [8] R. Dehdab, J.M. Brendel, S. Streich, et al., Evaluation of a deep learning denoising algorithm for dose reduction in whole-body photon-counting CT imaging: a cadaveric STUDY, *Acad. Radiol.* 32 (6) (2025) 3519–3531, <https://doi.org/10.1016/j.acra.2024.12.052>.
- [9] F. Faul, E. Erdfelder, A. Buchner, A.-G. Lang, Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses, *Behav. Res. Methods* 41 (4) (2009) 1149–1160, <https://doi.org/10.3758/BRM.41.4.1149>.
- [10] D.C. Benz, G. Benetos, G. Rampidis, et al., Validation of deep-learning image reconstruction for coronary computed tomography angiography: Impact on noise, image quality and diagnostic accuracy, *J. Cardiovasc. Comput. Tomogr.* 14 (5) (2020) 444–451, <https://doi.org/10.1016/j.jcct.2020.01.002>.
- [11] H. Wang, R. Wang, Y. Li, et al., Assessment of image quality of coronary computed tomography angiography in obese patients by comparing deep learning image reconstruction with adaptive statistical iterative reconstruction Veo, *J. Comput. Assist. Tomogr.* 46 (1) (2022) 34–40, <https://doi.org/10.1097/RCT.0000000000001252>.
- [12] J. Mei, C. Chen, R. Liu, H. Ma, Effect of deep learning image reconstruction on image quality and pericoronary fat attenuation index, *J. Imaging Inform Med.* 38 (3) (2025) 1881–1890, <https://doi.org/10.1007/s10278-024-01234-3>.
- [13] F. Tatsugami, T. Higaki, Y. Nakamura, et al., Deep learning-based image restoration algorithm for coronary CT angiography, *Eur. Radiol.* 29 (10) (2019) 5322–5329, <https://doi.org/10.1007/s00330-019-06183-y>.
- [14] P.-H. Jeon, S.-H. Jeon, D. Ko, et al., Assessment of image quality of coronary CT angiography using deep learning-based CT reconstruction, *Phantom and Patient Studies. Diagnostics (Basel)* 13 (11) (2023), <https://doi.org/10.3390/diagnostics13111862>.
- [15] J.H. Hong, E.A. Park, W. Lee, C. Ahn, J.H. Kim, Incremental image noise reduction in coronary CT angiography using a deep learning-based technique with iterative reconstruction, *Korean J. Radiol.* 21 (10) (2020) 1165–1177, <https://doi.org/10.3348/kjr.2020.0020>.
- [16] Y. Nagayama, T. Emoto, Y. Kato, et al., Improving image quality with super-resolution deep-learning-based reconstruction in coronary CT angiography, *Eur. Radiol.* 33 (12) (2023) 8488–8500, <https://doi.org/10.1007/s00330-023-09888-3>.
- [17] Y. Funama, T. Nakaura, A. Hasegawa, et al., Noise power spectrum properties of deep learning-based reconstruction and iterative reconstruction algorithms: phantom and clinical study, *Eur. J. Radiol.* 165 (2023) 110914, <https://doi.org/10.1016/j.ejrad.2023.110914>.
- [18] J. Osoria-Velasquez, G. Tremamunno, T. Emrich, et al., Deep Learning-based denoising improves myocardial late enhancement imaging with spectral photon-

- counting CT, *J. Cardiovasc. Comput. Tomogr.* (2025), <https://doi.org/10.1016/j.jcct.2025.12.002>.
- [19] L. Zhu, R. Ha, H. Machida, et al., Image quality of coronary CT angiography at ultra low tube voltage reconstructed with a deep-learning image reconstruction algorithm in patients of different weight, *Quant. Imaging Med. Surg.* 13 (6) (2023) 3891–3901, <https://doi.org/10.21037/qims-22-1141>.
- [20] A.K. Hara, R.G. Paden, A.C. Silva, J.L. Kujak, H.J. Lawder, W. Pavlicek, Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study, *AJR Am. J. Roentgenol.* 193 (3) (2009) 764–771, <https://doi.org/10.2214/AJR.09.2397>.
- [21] M.O. Chandran, S. Pendem, P.S. P. C. Chacko, R. Priyanka, Kadavigere, Comparison of image quality between deep learning image reconstruction and Iterative reconstruction technique for CT Brain- a pilot study, *F1000Res* 13 (2024) 691, <https://doi.org/10.12688/f1000research.150773.1>.
- [22] A. Delabie, R. Bouzerar, R. Pichois, X. Desdoit, J. Vial, C. Renard, Diagnostic performance and image quality of deep learning image reconstruction (DLIR) on unenhanced low-dose abdominal CT for urolithiasis, *Acta Radiol.* 63 (9) (2022) 1283–1292, <https://doi.org/10.1177/02841851211035896>.
- [23] K. Nasir, M. Cainzos-Achirica, Role of coronary artery calcium score in the primary prevention of cardiovascular disease, *BMJ* 373 (2021) n776, <https://doi.org/10.1136/bmj.n776>.
- [24] K.K. Wong, J.S. Cummock, Y. He, R. Ghosh, J.J. Volpi, S.T.C. Wong, Retrospective study of deep learning to reduce noise in non-contrast head CT images, *Comput. Med. Imaging Graph.* 94 (2021) 101996, <https://doi.org/10.1016/j.compmedimag.2021.101996>.
- [25] D. Plajer, M. Hahn, M. Chaika, et al., Deep-learning denoising minimizes radiation exposure in neck CT beyond the limits of conventional reconstruction, *Eur. J. Radiol.* 178 (2024) 111523, <https://doi.org/10.1016/j.ejrad.2024.111523>.
- [26] L.e. Cao, X. Liu, T. Qu, et al., Improving spatial resolution and diagnostic confidence with thinner slice and deep learning image reconstruction in contrast-enhanced abdominal CT, *Eur. Radiol.* 33 (3) (2023) 1603–1611, <https://doi.org/10.1007/s00330-022-09146-y>.
- [27] T. Kobayashi, T. Nishii, K. Umehara, et al., Deep learning-based noise reduction for coronary CT angiography: using four-dimensional noise-reduction images as the ground truth, *Acta Radiol.* 64 (5) (2023) 1831–1840, <https://doi.org/10.1177/02841851221141656>.
- [28] M. Shotwell, Diagnostic confidence in coronary artery disease characterization using a Ct deep learning reconstruction algorithm, *J. Cardiovasc. Comput. Tomogr.* 17 (4) (2023) S78–S79, <https://doi.org/10.1016/j.jcct.2023.05.194>.
- [29] D. Han, A. Shanbhag, R.J. Miller, et al., AI-derived left ventricular mass from noncontrast cardiac CT: correlation with Contrast CT angiography and CMR, *JACC Adv.* 3 (10) (2024) 101249, <https://doi.org/10.1016/j.jacadv.2024.101249>.