Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-based opportunistic coronary artery calcium assessment on non-gated chest CT: a systematic review and meta-analysis

Yu-Chuan Chen
National Yang Ming Chiao Tung University
Yun-Chu Wang
National Yang Ming Chiao Tung University
Marly van Assen
Emory University
Chih-Yu Yang
cyyang3@vghtpe.gov.tw

Taipei Veterans General Hospital

Article

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11485986/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. M.v.A. reports research funding from Siemens Healthineers, Elucid, Lucentia, and Cleerly, and serves as an advisor to Lucentia, and declares no non-financial competing interests. All other authors declare no financial or non-financial competing interests.

AI-based opportunistic coronary artery calcium assessment on non-gated chest CT: a systematic review and meta-analysis

Yu-Chuan Chen[1], Yun-Chu Wang[1], Marly van Assen[2], Chih-Yu Yang[1,3,4,5,6,*]

[1] Institute of Clinical Medicine, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[2] Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA, USA

[3] Division of Nephrology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[4] Stem Cell Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

[5] Center for Intelligent Drug Systems and Smart Bio-devices (IDS2B), National Yang Ming Chiao Tung University, Hsinchu, Taiwan

[6] International Medical Service Center, Taipei Veterans General Hospital, Taipei, Taiwan

*Corresponding author: Chih-Yu Yang, Department of Medicine, Taipei Veterans General Hospital, No. 201, Section 2, Shipai Road, Beitou District, Taipei 112201, Taiwan. Email: cyyang3@vghtpe.gov.tw.

ORCID: 0000-0001-9899-3159

Abstract

Coronary artery calcium (CAC) on routine non-gated chest CT can be quantified by artificial intelligence (AI), but performance at treatment-relevant thresholds and effects on preventive care are uncertain. We systematically reviewed 110 reports of AI-based CAC assessment on non-gated chest CT; 59 contributed to meta-analyses. Against paired ECG-gated CT, pooled sensitivity and specificity were 87.1% and 92.3% for CAC >0 and 83.1% and 96.3% for CAC ≥100, with specificity above 97% at ≥300 and ≥400. Category disagreement was mainly underestimation: 30.7% of gated CAC 1–99 was classified as 0 and 40.3% of CAC 100–299 below 100. Adjusted hazard ratios for major adverse cardiovascular events rose from 1.60 (CAC 1–100) to 3.39 (>400) versus CAC 0. In two randomised trials, AI-enabled notification increased six-month statin prescribing (51.2% vs 6.9% and 44.0% vs 9.8%). Substantial AI-detected CAC should prompt review of preventive treatment, whereas low AI-derived categories can underestimate disease burden.

Introduction

Cardiovascular disease remains a major cause of death and disability, making timely identification and treatment of people at increased risk a priority for prevention.[1] Coronary artery calcium (CAC) is an imaging marker of coronary atherosclerotic burden that can inform preventive treatment decisions alongside clinical risk assessment.[2,3] Dedicated electrocardiogram (ECG)-gated CT synchronises image acquisition with the cardiac cycle to reduce motion artefacts and is the established method for CAC quantification. However, it requires a separate examination, with costs and access barriers that limit its use.[4,5] For patients already undergoing chest CT for other clinical indications, assessment of CAC on these non-ECG-gated images (hereafter non-gated CT) offers an opportunity to identify coronary atherosclerosis without an additional scan and extend cardiovascular prevention to patients being investigated for other conditions.[5,6]

Artificial intelligence (AI) can support this approach by identifying and quantifying CAC on routinely acquired CT images. Models provide numerical calcium scores, severity categories, or both, enabling assessment across large numbers of examinations without requiring clinicians to measure calcium individually.[5,7] Commercial systems have received regulatory clearance,[5] and the 2026 US dyslipidaemia guideline explicitly incorporates incidental CAC assessed visually or with a validated AI algorithm into decisions about lipid-lowering therapy. Its recommendations link the extent of coronary atherosclerosis to treatment intensity and lipid goals, giving AI-derived assessments a concrete role in preventive care.[3]

As AI-based opportunistic CAC assessment enters clinical practice, its performance needs to be considered in relation to the preventive decisions it is intended to inform. A previous meta-analysis of 25 studies reported strong but heterogeneous agreement between AI and manual CAC scoring across studies that included dedicated cardiac CT as well as non-gated examinations,[8] and two scoping reviews described prognostic associations and early clinical implementation without meta-analysis; one was limited to English-language studies and the other used single-reviewer screening without formal quality assessment.[9,10] However, high overall agreement can coexist with

misclassification at CAC thresholds that influence whether lipid-lowering treatment is started or intensified. The evidence has also expanded beyond measurement and risk prediction: recent studies have evaluated whether incorporating AI-detected CAC into clinical care changes preventive management.[11,12] An updated synthesis focused on non-gated CT is therefore needed to assess performance at treatment-relevant levels of disease burden alongside this emerging evidence.

We therefore conducted a systematic review and meta-analysis of AI-based opportunistic CAC assessment on non-gated CT to evaluate diagnostic accuracy and severity classification, associations with subsequent cardiovascular events, and the effects of AI-enabled care pathways on preventive management.

Results

Study selection and characteristics

Searches identified 6,542 records, of which 3,308 were removed before screening (2,467 duplicates, 795 conference abstracts, and 46 registrations or superseded versions). Of the 3,234 titles and abstracts screened, 341 reports were sought, seven could not be retrieved, and 334 full texts were assessed. Overall, 110 reports were included,[7,11–119] of which 59 contributed to at least one meta-analysis (Fig. 1; Table 1; Supplementary Tables 19 and 20).

Characteristics of all included reports are presented in Supplementary Table 5A. By principal evidence contribution, 32 reports addressed the paired ECG-gated reference, 37 the same-scan manual reference, nine the expert visual reference, 24 prognosis, and four implementation pathways, and four provided other eligible evidence (scan–rescan repeatability, a radiation-dose comparison, and accuracy data that could not be pooled); 51 reports evaluated commercial systems and 56 academic or research algorithms. Studies included lung-screening, routine chest CT, oncology, and PET-CT attenuation-correction settings, with scanners from multiple manufacturers, section thicknesses from submillimetre to 5 mm, and soft-tissue and sharp kernels. Overall risk of bias was high in 40 (51%) of 78 paired-gated and 36 (49%) of 73 same-scan manual diagnostic estimates; of 16 major adverse cardiovascular event (MACE) estimates, 11 were at moderate and five at high risk of bias (Fig. 2; Supplementary Fig. 1; Supplementary Tables 3A–4).

Pooled sensitivity and specificity by reference standard and threshold

Sensitivity and specificity were pooled separately for each reference standard and threshold (CAC >0, ≥100, ≥300, and ≥400 Agatston units [AU]); the paired ECG-gated comparisons at CAC >0 and ≥100 were the co-primary outcomes (Table 2). Against paired ECG-gated CT, pooled sensitivity for CAC >0 was 87.1% (95% confidence interval [CI] 83.2–90.2) and specificity was 92.3% (87.2–95.4), based on 24 estimates from 17 reports (n = 9,393). At CAC ≥100, 26 estimates from 19 reports (n = 10,940) yielded sensitivity of 83.1% (78.5–86.8) and specificity of 96.3% (94.2–97.7). Sensitivity and specificity were 80.5% (61.5–91.5) and 98.9% (96.2–99.7) at CAC ≥300, and 80.1% (74.3–84.9) and 97.7% (96.2–98.7) at CAC ≥400 (Figs. 3 and 4; Supplementary Table 6; Supplementary Fig. 2).

Against manual scoring on the same non-gated image, pooled sensitivities at CAC >0, ≥100, ≥300, and ≥400 were 93.8%, 93.3%, 90.5%, and 90.3%, respectively; corresponding specificities were 88.7%, 97.5%, 98.5%, and 98.7% (Fig. 5; Fig. 6a, b; Supplementary Table 6). Against expert visual assessment, sensitivity for detecting any CAC was 95.8% (95% CI 93.9–97.1) and specificity was 90.3% (68.0–97.6), based on four reports (n = 1,043; Fig. 6c). Comparable data supported pooling of CAC presence only; visual severity grades were summarised descriptively.

Agreement of AI-derived CAC categories with ECG-gated CT

For patients in each ECG-gated CAC category, cross-classification analyses estimated how AI on the paired non-gated CT categorised them: the probability of the same category (correct classification), a lower category (underestimation), or a higher category (overestimation). Six cross-classification tables from four reports (n = 1,727) informed the primary analysis using ECG-gated categories of 0, 1–99, 100–299, and ≥300 AU (Fig. 7a). Probabilities of correct classification were 95.5%, 62.5%, 51.9%, and 78.3%, respectively, and underestimation was more frequent than overestimation in both intermediate categories. For gated CAC 1–99, the probability of classification as 0 was 30.7% (95% credible interval [CrI] 13.8–54.1), compared with 6.8% for classification as ≥100. For gated CAC 100–299, the probability of classification below 100 was 40.3% (17.2–69.6), compared with 7.7% (1.4–22.5) for classification as ≥300. Among those with gated CAC ≥300, 21.7% (8.9–43.0) were classified below 300 and 4.1% (1.0–11.8) below 100. No gated CAC 0 case was classified as ≥100 in the contributing tables (model-based probability 0.6% [0.1–1.7]; Supplementary Table 8; Supplementary Fig. 4).

The alternative scheme of 0, 1–99, 100–399, and ≥400 was analysed separately using 20 tables from 15 reports (n = 7,935), none shared with the primary analysis (Fig. 7b). The probability of classifying gated CAC 1–99 as 0 was 30.7% (95% CrI 20.7–41.9). For gated CAC 100–399, 31.9% (23.2–41.0) were classified below 100 and 6.5% (3.7–10.3) as ≥400. Among those with gated CAC ≥400, 19.5% (15.0–24.7) were classified below 400 (complete probabilities, leave-one-report-out analyses, and weighted κ in Supplementary Tables 9 and 10 and Supplementary Fig. 5).

Sensitivity analyses, publication bias, and meta-regression

Restriction to estimates at low overall risk of bias yielded sensitivity and specificity of 87.5% and 84.8% at CAC >0 and 78.5% and 94.8% at CAC ≥100; restriction to confirmed non-contrast CT yielded 86.7% and 92.4%, and 82.2% and 96.5%, respectively (Supplementary Table 6). Deeks regression suggested funnel asymmetry at CAC ≥100 (P = 0.067) and ≥400 (P = 0.054) but not at CAC >0 (P = 0.14); with four reports at CAC ≥300, asymmetry testing was of limited value (Supplementary Fig. 3). Certainty for paired-gated diagnostic outcomes was low or very low (Supplementary Table 7).

After false-discovery-rate (FDR) correction, performance differed by commercial status at CAC >0, ≥100, and ≥400, predominantly in specificity; at CAC ≥100, sensitivity was 81.6% for commercial and 83.9% for non-commercial systems, and specificity was 98.5% and 88.7% (global q < 0.0001). Publication-year band was associated with performance at CAC >0, and country or region and model

architecture at CAC ≥400; no other covariate was associated with performance where estimable, and none met the minimum information requirement at CAC ≥300 (Supplementary Table 13; Supplementary Figs. 7 and 8; Supplementary Data 1, sheet S6).

Agreement of Agatston scores and sources of discrepancy

Twenty-six reports provided 39 Bland–Altman comparisons of whole-heart Agatston scores (25 against paired ECG-gated CT and 14 against same-scan manual scoring). Limits of agreement varied substantially, and several comparisons showed mean biases close to zero despite 95% limits of agreement extending several hundred AU in either direction (Supplementary Fig. 6). No report derived non-gated-specific CAC cutpoints by calibration against paired ECG-gated categories and then validated the full set of cutpoints in an independent sample (Supplementary Data 1, sheet S5).

In six reports comparing human and AI assessments of the same non-gated images against a common gated reference, both underestimated CAC categories in the most comparable studies, although none formally tested human-versus-AI superiority (Supplementary Tables 11 and 12). With the scoring model held constant, exact category agreement was 91.0%, 84.8%, and 62.5% when the same raw CT data were reconstructed at 1, 3, and 5 mm.[40] When the same algorithm scored both examinations in the same 652 patients, category accuracy against the gated manual reference was 96.8% on the gated CT but 86.0% on the denoised non-gated CT.[24] Deep-learning conversion of kernel and slice thickness increased the concordance correlation coefficient from 0.269 to 0.881 and weighted κ from 0.115 to 0.792.[23]

AI-derived CAC and major adverse cardiovascular events

Higher AI-derived CAC categories were associated with subsequent MACE (Fig. 8). Compared with CAC 0, pooled adjusted hazard ratios were 1.60 (95% CI 1.29–1.97; six reports) for CAC 1–100, 2.40 (1.49–3.89; five reports) for CAC 101–400, and 3.39 (1.65–6.98; four reports) for CAC >400. Between-study heterogeneity increased across these categories (I^2 = 25.5%, 71.0%, and 78.1%, respectively), with corresponding 95% prediction intervals of 1.12–2.27, 0.94–6.15, and 1.01–11.42. Certainty was moderate for CAC 1–100 and low for the higher categories (Supplementary Tables 14–16).

AI-enabled notification pathways and preventive treatment

Two randomised trials evaluated AI-enabled CAC notification pathways (Table 3; Supplementary Tables 17 and 18). In NOTIFY-1, statins were prescribed within six months to 44 of 86 participants (51.2%) assigned to notification and six of 87 (6.9%) receiving usual care (risk ratio [RR] 7.42, 95% CI 3.34–16.50).[11] In NOTIFY-PICTURE, the corresponding proportions were 44 of 100 (44.0%) and ten of 102 (9.8%; RR 4.49, 2.39–8.42).[12] A secondary fixed-effect synthesis yielded RR 5.44 (3.32–8.92), and certainty was high for the effect of the complete pathways on six-month prescribing.

In NOTIFY-PICTURE, CAC had been mentioned in 196 (97.0%) of 202 original radiology reports but in only eight (4.0%) report impressions. Notification increased clinical encounters and downstream testing in NOTIFY-1, including coronary artery disease testing (13 of 86 vs two of 87); NOTIFY-

PICTURE reported cardiovascular testing in 41 of 100 versus 28 of 102 participants and no increase in health-care utilisation. Among NOTIFY-1 participants with follow-up lipid measurements (50 vs 29), mean low-density lipoprotein (LDL) cholesterol was 97.2 mg/dL (SD 30.3) versus 115.3 mg/dL (29.4; P = 0.005).

In NOTIFY-EXTEND, uncontrolled follow-up of early- and delayed-notification recipients from NOTIFY-1 showed a reduction in mean LDL cholesterol from 115.6 mg/dL (SD 28.2) to 90.6 mg/dL (35.1; P < 0.01) among 97 participants retested at 12 months. Of 86 participants who initiated statins, 75 (87%) remained treated at 18 months.[106] Neither randomised trial assessed effects on cardiovascular events.

Discussion

In this systematic review and meta-analysis, AI-based CAC assessment on non-gated CT showed high specificity against paired ECG-gated CT at thresholds used to guide prevention, although sensitivity was lower and severity classification tended towards underestimation. Higher AI-derived CAC was associated with greater subsequent MACE risk, and randomised trials showed that incorporating AI-detected CAC into notification pathways substantially increased statin prescribing. AI-CAC can therefore extend assessment of coronary atherosclerosis to patients undergoing CT for other indications, with its clinical value depending on how the findings inform subsequent care.

The opportunity extends beyond patients whose cardiovascular risk has been underestimated. Among 5,678 adults without known atherosclerotic cardiovascular disease (ASCVD) undergoing routine non-gated CT, those with AI-CAC ≥100 had an average estimated 10-year ASCVD risk of approximately 24%, yet only 26% were receiving statins.[120] Recent discussions of disease-based prevention have emphasised recognising and treating subclinical atherosclerosis, and reviews of incidental CAC have described how existing CT examinations could support this approach.[121,122] Our findings provide AI-specific evidence for bringing it into routine imaging, potentially reaching patients with unrecognised disease and those whose recognised risk has not led to treatment. Using these findings in treatment discussions may improve adherence and risk-factor control.[123,124]

High specificity against paired ECG-gated CT at CAC ≥100 and higher thresholds supports using substantial AI-detected CAC to prompt preventive treatment review, whereas lower sensitivity cautions against interpreting a low AI-derived score as evidence that clinically important disease is absent. In the intermediate categories, AI more often underestimated than overestimated the category, with substantial uncertainty, and a larger, separate evidence base using the 400-AU scheme showed a similar pattern. A previous meta-analysis reported strong agreement between AI and manual scoring (κ 0.83) across gated and non-gated CT, and a scoping review described AI misclassification as mostly underestimation without pooling estimates;[8,9] our threshold-specific and cross-classification analyses quantify where, and in which direction, disagreement occurs. Classification below 100 could weaken the indication for intensive lipid lowering. Movement between CAC 100–299 and 300–999 might leave the initial LDL-cholesterol goal unchanged, although the higher category can justify further intensification.[3] Category disagreement should therefore be

interpreted in relation to the treatment decision it might change, and an AI report of no CAC should not overturn an otherwise established indication for lipid-lowering therapy.

Some discrepancy appears to arise from the images themselves: human readers of the same non-gated scans also underestimated categories in the most comparable studies, and performance changed when reconstruction settings varied with the AI scorer held constant.[23,40] Accurate scoring of a non-gated image therefore need not reproduce the burden measured on a dedicated gated examination. If the AI-reported category conflicts with calcification visible on the images, or would change treatment, the images should be reviewed; a dedicated gated CT is warranted only when its result would alter management. Future studies should test whether image harmonisation reduces underestimation while preserving specificity at treatment thresholds.

The prognostic evidence supports retaining severity information: adjusted MACE hazard ratios increased across AI-derived categories, although heterogeneity was greater at higher burdens. This is consistent with Multi-Ethnic Study of Atherosclerosis (MESA) evidence of similar prognostic value for manually assessed CAC on non-gated and gated examinations.[125] Useful risk stratification can therefore persist despite imperfect numerical correspondence with gated scores, although these associations do not establish that individual AI-derived and gated scores are interchangeable, or determine absolute treatment benefit for a particular patient.

The notification trials show that this information can change preventive care, with substantial increases in prescribing of a treatment with established cardiovascular benefit.[11,12] The interventions combined AI detection with communication and clinical follow-up; their effects support the complete intervention but do not establish the incremental benefit of AI over other methods of identifying CAC. Because CAC had already been mentioned in 97% of original radiology reports in NOTIFY-PICTURE, the trial illustrates a failure of reported disease to prompt treatment as well as an opportunity to improve detection. Together, the trials support actively communicating clinically important CAC to the clinician responsible for prevention, with information that helps patients understand why treatment is being considered.

The next objective should be appropriate and sustained treatment after disease is identified. The 2026 dyslipidaemia guideline supports high-intensity statin therapy for moderate-to-severe incidental CAC assessed visually or with a validated AI algorithm, targeting at least a 50% LDL-cholesterol reduction and a concentration below 70 mg/dL.[3] In CAUGHT-CAD, disclosure of gated CAC with education and statin treatment produced greater LDL-cholesterol reduction and slower plaque progression over three years;[126] although it did not evaluate opportunistic AI-CAC, it provides a model for what should follow detection. NOTIFY-EXTEND suggests treatment persistence, although its follow-up was uncontrolled.[106] Future trials should assess sustained LDL-cholesterol control, including in patients whose existing lipid-lowering therapy might be insufficient for their disease burden.

Implementation should distinguish preventive management from investigation of possible symptomatic disease. For asymptomatic patients, incidental CAC generally calls for risk-factor assessment and treatment rather than routine additional coronary imaging or ischaemia testing.[127] A

useful AI-CAC service would make the burden clear in the radiology report and connect the finding to a clinician able to review symptoms and current treatment. Evaluations should measure treatment uptake and lipid control alongside the appropriateness of additional testing and the workload of follow-up, particularly in patients undergoing CT for pulmonary or oncological disease, in whom competing illness and preferences influence the value of prevention.

This review brings together evidence on measurement, prognosis, and preventive care while distinguishing reference standards that address different aspects of performance. Several limitations affect interpretation. Around half of the selected diagnostic estimates were at high risk of bias, and certainty for paired-gated diagnostic outcomes was low or very low. The primary cross-classification analysis was based on four reports, although a larger, non-overlapping analysis supported a similar pattern. Evidence against expert visual assessment supported pooling of CAC presence only, and most quantitative validation data came from non-contrast CT. Study-level comparisons could not reliably separate algorithm effects from acquisition protocols and populations, and prognostic studies differed in adjustment and setting. Most reports came from North America, Europe, and east Asia (Supplementary Table 5A), which limits assessment of applicability to other regions and populations. The two randomised implementation trials were conducted at single centres and did not assess cardiovascular events.

AI-based opportunistic CAC assessment provides an opportunity to act on coronary atherosclerosis that would otherwise remain outside, or fail to influence, preventive care. The evidence supports its use to identify substantial disease and inform treatment, while recognising that AI can underestimate severity categories; multicentre trials should establish whether improved prescribing is sustained and leads to effective lipid control and fewer cardiovascular events.

Methods

Protocol and registration

This systematic review and meta-analysis was registered prospectively with INPLASY (INPLASY202660006) and is reported according to PRISMA 2020, PRISMA-DTA, and PRISMA-S (Supplementary Section 9).[128–130] Deviations from the registered protocol are listed in Supplementary Table 1.

Search strategy

We searched PubMed, Embase, Web of Science Core Collection, and the Cochrane Central Register of Controlled Trials from inception to 2 October 2026, without date or language restrictions, using terms for CAC, non-gated thoracic CT, and artificial intelligence, including machine learning and deep learning; a supplementary search targeted implementation and preventive-care reports that might not mention AI in the title or abstract (full strategies in Supplementary Note 1). Records were deduplicated by DOI, PMID, and normalised title, and conference abstracts, registrations, and superseded versions were removed before screening. Study authors were not contacted.

Eligibility criteria

Eligible reports enrolled adults undergoing routine non-gated chest CT, including low-dose lung-screening, diagnostic, oncology, radiotherapy-planning, and PET/CT attenuation-correction acquisitions, and evaluated an automated, machine-learning, or deep-learning method for CAC detection, quantification, categorisation, or prognostic assessment, or a care pathway incorporating AI-detected CAC. Eligible reference standards were quantitative assessment on paired ECG-gated calcium-scoring CT, manual or semi-manual Agatston assessment on the same non-gated image, and expert visual CAC grading. Diagnostic accuracy, agreement, cohort, prognostic, randomised, non-randomised, before–after, and quality-improvement designs were eligible when a relevant outcome was extractable, as were full-length conference proceedings and preprints with sufficient methods. We excluded abstract-only records, duplicate or superseded versions, gated-only or coronary CT angiography-only studies without separately reported non-gated data, manual-only assessments, phantom, simulation, or animal studies, series with fewer than 10 participants, reviews, studies of non-coronary calcification, and reports without extractable eligible outcomes. When both uncorrected and manually corrected AI results were reported, the uncorrected output was prioritised.

Study selection and data extraction

Two reviewers (Y.-C.C. and Y.-C.W.) independently screened titles and abstracts, assessed full texts, and extracted study, participant (including sex), imaging, algorithm, reference-standard, and outcome data; disagreements were resolved by consensus or by a third reviewer (C.-Y.Y.). Reports not retrieved and reports excluded after full-text assessment, with reasons, are listed in Supplementary Tables 19 and 20. Calcium-volume and mass measurements were not converted to Agatston units (AU). Threshold-specific 2×2 tables were reconstructed from reported category tables where necessary, and thresholds defined as >100 or >400 were grouped with ≥100 and ≥400. Throughout, an estimate denotes a threshold-specific result from one participant cohort against one reference standard. Reports were linked by named datasets, institutions, enrolment periods, sample sizes, and algorithm families; within each synthesis, cohorts judged to overlap definitely or probably were represented once, and possible overlap was documented (Supplementary Table 5B; Supplementary Data 1, sheet S12). When several outputs were available for the same participants, we retained the prespecified or report-designated primary uncorrected output, then prioritised external validation and the most complete cohort, without reference to observed performance.

Reference standards and outcomes

Three reference standards were analysed separately: paired ECG-gated calcium-scoring CT, which assessed whether AI reproduced dedicated calcium-scoring classifications; manual quantitative assessment of the same non-gated image, which compared AI and human scoring of identical images; and expert visual assessment (Supplementary Table 2). The prespecified co-primary outcomes were sensitivity and specificity of AI-derived CAC at thresholds of >0 and ≥100 AU against paired ECG-gated CT; ≥300 and ≥400 AU were secondary thresholds. CAC >0 identifies calcified coronary atherosclerosis, CAC ≥100 supports intensive lipid lowering, and CAC ≥300 identifies a

burden for which current guidance supports considering further intensification;[3] 400 AU defines the highest category of an established alternative classification. For category agreement, we estimated the probability of each AI-derived category within each ECG-gated reference category, separately for the primary scheme (0, 1–99, 100–299, and ≥300 AU) and the alternative scheme (0, 1–99, 100–399, and ≥400 AU), without converting observations between schemes. Continuous-score agreement (mean bias and Bland–Altman limits of agreement) was summarised descriptively by reference standard. For prognosis, we examined adjusted Cox-model hazard ratios for study-defined MACE, comparing AI-derived CAC 1–100, 101–400, and >400 with CAC 0. The primary implementation outcome was statin prescribing within six months of assignment to an AI-enabled CAC care pathway versus usual care; treatment persistence, lipid outcomes, downstream testing, health-care use, and cardiovascular events were also examined.

Risk of bias and applicability

Diagnostic-accuracy estimates were assessed with QUADAS-3 (version 1.2), prognostic estimates with QUIPS, and randomised trials with RoB 2.[131–133] The unit of diagnostic assessment was the selected estimate, defined by report, cohort, algorithm, reference standard, threshold, analysis population, and synthesis question; risk of bias was judged for the Participants, Index test, Target condition, and Analysis domains and applicability for the first three, with overall judgments of Low, High, or Insufficient information (High if any domain was High; Low if all were Low). The 20 official QUADAS-3 signalling questions were retained verbatim, with one additional, separately labelled question on independence between AI development and evaluation data. Two reviewers assessed each estimate independently against the full report and supplementary material, with adjudication by C.-Y.Y. when required; all judgments, rationales, and signalling-question responses are provided in Supplementary Data 1 (sheets S2–S4).

Certainty of evidence

Certainty was assessed post hoc with GRADE for the paired-gated diagnostic outcomes at CAC >0, ≥100, and ≥300 (sensitivity and specificity rated separately), for each MACE contrast, and for six-month statin prescribing after an AI-enabled notification pathway, using the GRADE frameworks for test accuracy, prognostic factors, and interventions, respectively.[134–140] Domain-specific judgments and their rationale are given in Supplementary Tables 7, 16, and 18; the implementation estimand was assignment to the complete pathway rather than the isolated effect of AI.

Statistics and reproducibility

Diagnostic accuracy. Threshold-specific 2×2 tables were pooled with bivariate random-effects models (Reitsma parameterisation, restricted maximum likelihood [REML]), fitted separately for each reference standard and threshold.[141,142] A continuity correction of 0.5 was applied only to individual estimates containing a zero cell; individual estimates are shown with exact binomial 95% CIs and pooled estimates with model-based 95% CIs and prediction intervals. Summary receiver operating characteristic curves used the Rutter–Gatsonis form of each model, and likelihood ratios and predictive values at assumed prevalences of 10% to 40% were derived from the joint model

covariance (Supplementary Data 1, sheets S9B and S9C). Sensitivity analyses refitted the paired-gated models with one contribution per participant cohort after report linkage and, at CAC >0 and ≥100, after restriction to estimates at low risk of bias, aggregation of non-overlapping cohorts within reports, restriction to confirmed non-contrast CT, and two extreme allocations of failed or unassessable scans (Supplementary Table 6; Supplementary Data 1, sheets S9A and S9D). Small-study effects were assessed with threshold-specific Deeks weighted regression after dependent estimates were aggregated within reports; a two-sided P value below 0.10 indicated funnel asymmetry, and analyses with fewer than ten reports were interpreted descriptively.[143]

Category agreement. Report- and cohort-level 4×4 cross-classification tables were analysed with a Bayesian hierarchical multinomial logistic-normal model, an extension of multivariate random-effects approaches for staging tables and multiple-threshold tests.[144–146] Within each ECG-gated reference category, the correct-classification cell was the baseline and the remaining logits varied across tables, with Normal(0, 2) priors on pooled logit means and half-Normal(0, 1) priors on between-table standard deviations. Models were fitted in Stan with four chains of 2,000 warm-up and 2,000 sampling iterations,[147] and posterior means with equal-tailed 95% credible intervals (CrIs) are reported; probabilities of underestimation and overestimation were derived within each posterior draw. Leave-one-report-out analyses omitted all tables associated with each report, including related reports sharing participants. Convergence criteria and diagnostics are reported in Supplementary Table 8. Linear-weighted κ was calculated from complete tables, with variances from 10,000 multinomial bootstrap samples, and pooled with REML random-effects models and modified Hartung–Knapp confidence intervals (Supplementary Table 10).

Continuous agreement and sources of discrepancy. Bland–Altman bias and limits of agreement on the native Agatston scale were summarised descriptively because imaging protocols and score distributions differed between studies; reference-minus-index definitions were reoriented to index minus reference. Human-versus-AI comparisons were restricted to reports in which both assessed the same non-gated image against the same paired ECG-gated reference and were summarised descriptively, because separate index-versus-reference tables do not preserve the paired discordance required for a formal superiority test. Studies comparing different image inputs with the same algorithm were grouped by what was held constant (Supplementary Tables 11 and 12).

Subgroup analyses and meta-regression. Prespecified study-level covariates (commercial status, country or region, CT clinical setting, model architecture, paired-scan timing, publication year, reconstruction kernel, scanner vendor, slice thickness, training ground-truth pairing, and validation design) were examined at each paired-gated threshold by bivariate meta-regression jointly modelling logit sensitivity and logit false-positive rate; contrast status was summarised descriptively. Formal fitting required at least 10 contributing reports and at least two levels, each with at least three estimates from three reports. Global effects were assessed with likelihood-ratio tests and Benjamini–Hochberg false-discovery-rate adjustment within each threshold;[148] component coefficients were used only to describe the direction of an FDR-significant result, and absolute subgroup estimates were

obtained from subgroup-specific bivariate models (Supplementary Table 13; Supplementary Figs. 7 and 8).

Prognosis. The prespecified fully adjusted hazard ratio with a complete 95% CI was selected from each report, and adjusted hazard ratios were pooled on the log scale with REML random-effects models and Hartung–Knapp inference, with 95% prediction intervals where at least three reports contributed.[149,150] One proximal-CAC estimate was retained for narrative synthesis because its contrast was not exchangeable with the harmonised categories.

Implementation. The two randomised trials were presented individually because their bundled workflows differed; six-month statin prescribing was expressed as risk ratios with 95% CIs, with a secondary fixed-effect inverse-variance synthesis that was not interpreted as the isolated effect of AI. Other outcomes and uncontrolled follow-up were summarised descriptively.

All statistical tests were two-sided; apart from the prespecified criteria for Deeks regression ($P < 0.10$) and FDR-adjusted $q < 0.05$, inference relied on estimates with 95% CIs or CrIs. Analyses used R version 4.5.3 with mada 0.5.12, metafor 5.0.1, and CmdStan 2.39.0 through cmdstanr 0.9.0.[151–154]

Ethics

This study synthesised aggregate data from published reports and did not involve human participants; ethical approval and consent were not required.

Use of large language models

A large language model (GPT 6 Astra, OpenAI) was used to improve the language and readability of the manuscript text; the authors reviewed and edited all output, verified its accuracy against the source data and references, and take full responsibility for the content.

Data availability

All analyses used aggregate data extracted from published reports; no individual participant data were obtained. The report-level characteristics and aggregate datasets supporting the analyses, including threshold-specific 2×2 data, cross-classification tables, complete subgroup and meta-regression results, and the source data underlying all figures and tables, are provided as Supplementary Data 1 with this Article. The registered protocol is available from INPLASY (INPLASY202660006; https://doi.org/10.37766/inplasy2026.6.0006).

Code availability

The R and Stan code used for the analyses is available from the corresponding author on reasonable request.

Acknowledgements

We thank Hao-Chen Chuang for assistance with the Embase search. This study received no funding.

Author contributions

Y.-C.C. and Y.-C.W. conceived the review with C.-Y.Y., and undertook study selection, data extraction, and assessments of risk of bias, applicability, and certainty; they directly accessed and verified all aggregate data reported in this Article. C.-Y.Y. provided adjudication when required, supervised the review, and is the guarantor. Y.-C.C. curated the data, did the statistical analyses, prepared the figures, and drafted the manuscript. M.v.A. provided guidance on manuscript writing and critically reviewed the manuscript. All authors interpreted the findings, critically revised the manuscript, had access to the aggregate data, and approved the submitted version.

Competing interests

M.v.A. reports research funding from Siemens Healthineers, Elucid, Lucentia, and Cleerly, and serves as an advisor to Lucentia, and declares no non-financial competing interests. All other authors declare no financial or non-financial competing interests.

References

1. Palaniappan, L. P. et al. 2026 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation* **153**, e275–e906 (2026).
2. Detrano, R. et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N. Engl. J. Med.* **358**, 1336–1345 (2008).
3. Blumenthal, R. S. et al. 2026 ACC/AHA/AACVPR/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of dyslipidemia: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **153**, e1154–e1276 (2026).
4. Agatston, A. S. et al. Quantification of coronary artery calcium using ultrafast computed tomography. *J. Am. Coll. Cardiol.* **15**, 827–832 (1990).
5. Foraker, R. et al. Opportunistic detection of coronary artery calcium on noncardiac chest computed tomography: an emerging tool for cardiovascular disease prevention: a scientific statement from the American Heart Association. *Circulation* **152**, e391–e401 (2025).
6. Hecht, H. S. et al. 2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: a report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology. *J. Thorac. Imaging* **32**, W54–W66 (2017).
7. Zeleznik, R. et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat. Commun.* **12**, 715 (2021).
8. Wang, T.-W. et al. Meta-analysis of deep learning approaches for automated coronary artery calcium scoring: performance and clinical utility. *Comput. Biol. Med.* **183**, 109295 (2024).
9. O'Toole, F. E. et al. Advances in artificial intelligence for non-ECG-gated coronary artery calcium scoring: a scoping review. *Cureus* **17**, e94019 (2025).
10. McDonald, V. J. F. et al. Prognostication and clinical opportunities with AI for coronary artery calcium: a scoping review. *BMJ Digit. Health* **2**, e000053 (2026).
11. Sandhu, A. T. et al. Incidental coronary artery calcium: opportunistic screening of previous nongated chest computed tomography scans to improve statin rates (NOTIFY-1 project). *Circulation* **147**, 703–714 (2023).
12. Dudum, R. et al. Effects of real-time notification of AI-detected incidental coronary artery calcium on statin prescription: the NOTIFY-PICTURE trial. *Circulation* **153**, 210–212 (2026).
13. Chen, Y. et al. Impact of scanning parameters on deep learning coronary artery calcium scoring in non-gated chest CT. *BMC Med. Imaging* **26**, 232 (2026).
14. Cho, S.-G. et al. Coronary artery calcium measurement on attenuation correction computed tomography using artificial intelligence: correlation with coronary flow capacity and prognosis. *Eur. J. Nucl. Med. Mol. Imaging* **52**, 1050–1059 (2025).
15. Dobrolinska, M. M. et al. Performance of visual, manual, and automatic coronary calcium scoring of cardiac 13N-ammonia PET/low dose CT. *J. Nucl. Cardiol.* **30**, 239–250 (2023).
16. Dobrolinska, M. M. et al. The prognostic value of visual and automatic coronary calcium scoring from low-dose computed tomography-[15O]-water positron emission tomography. *Eur. Heart J. Cardiovasc. Imaging* **25**, 1186–1196 (2024).
17. Eng, D. et al. Automated coronary calcium scoring using deep learning with multicenter external validation. *npj Digit. Med.* **4**, 88 (2021).

18. Gokmen, M. S. et al. A framework for cross-domain generalization in coronary artery calcium scoring across gated and non-gated computed tomography. Preprint at https://doi.org/10.48550/arXiv.2602.21935 (2026).
19. Hagopian, R. et al. AI opportunistic coronary calcium screening at Veterans Affairs hospitals. *NEJM AI* **2**, AIoa2400937 (2025).
20. Homayounieh, F. et al. Prediction of coronary calcification and stenosis: role of radiomics from low-dose CT. *Acad. Radiol.* **28**, 972–979 (2021).
21. Jung, J.-H., Lee, J. E., Lee, H. S., Yang, D. H. & Lee, J.-G. Diffusion-based image translation model from low-dose chest CT to calcium scoring CT with random point sampling. *Comput. Biol. Med.* **194**, 110506 (2025).
22. Kang, H. W. et al. Evaluation of fully automated commercial software for Agatston calcium scoring on non-ECG-gated low-dose chest CT with different slice thickness. *Eur. Radiol.* **33**, 1973–1981 (2023).
23. Kim, C. et al. Impact of deep learning-based image conversion on fully automated coronary artery calcium scoring using thin-slice, sharp-kernel, non-gated, low-dose chest CT scans: a multi-center study. *Korean J. Radiol.* **26**, 759–770 (2025).
24. Kim, S. et al. Performance of fully automated deep-learning-based coronary artery calcium scoring in ECG-gated calcium CT and non-gated low-dose chest CT. *Eur. Radiol.* **35**, 7084–7095 (2025).
25. Li, J. et al. Combating coronary calcium scoring bias for non-gated CT by semantic learning on gated CT. *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops* 2575–2583 (2023).
26. Lin, Y. et al. The role of artificial intelligence in coronary calcium scoring in standard cardiac computed tomography and chest computed tomography with different reconstruction kernels. *J. Thorac. Imaging* **39**, 111–118 (2024).
27. Ludwig, K. D. et al. AI-based coronary artery calcification on non-contrast CT: performance across calcium scoring, lung cancer screening, and liver transplant candidate cohorts. Preprint at https://doi.org/10.64898/2026.05.12.26352904 (2026).
28. Marjanski, T. et al. Can artificial intelligence help us in the evaluation of coronary artery calcification scores by acting as a prognosticator in patients that are operated on due to non-small cell lung cancer? A pivotal study. *J. Clin. Med.* **13**, 6579 (2024).
29. McKinney, S. E. et al. Evaluation of an artificial intelligence model for opportunistic Agatston scoring on non-gated chest computed tomography. *Sci. Rep.* **15**, 39535 (2025).
30. Morf, C. et al. Diagnostic value of fully automated artificial intelligence powered coronary artery calcium scoring from 18F-FDG PET/CT. *Diagnostics (Basel)* **12**, 1876 (2022).
31. Oh, H. S. et al. Feasibility and limitations of deep learning–based coronary calcium scoring in PET-CT: a comparison with coronary calcium score CT. *Eur. Radiol.* **34**, 4077–4088 (2024).
32. Pieszko, K. et al. Calcium scoring in low-dose ungated chest CT scans using convolutional long-short term memory networks. *Proc. SPIE* **12032**, 120323A (2022).
33. Pieszko, K. et al. Deep learning of coronary calcium scores from PET/CT attenuation maps accurately predicts adverse cardiovascular events. *JACC Cardiovasc. Imaging* **16**, 675–687 (2023).
34. Sartoretti, E. et al. Opportunistic deep learning powered calcium scoring in oncologic patients with very high coronary artery calcium (≥ 1000) undergoing 18F-FDG PET/CT. *Sci. Rep.* **12**, 19191 (2022).
35. van Assen, M. et al. Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: a validation study. *Eur. J. Radiol.* **134**, 109428 (2021).
36. van Velzen, S. G. M. et al. Automated cardiovascular risk categorization through AI-driven coronary calcium quantification in cardiac PET acquired attenuation correction CT. *J. Nucl. Cardiol.* **30**, 955–969 (2023).
37. Xu, J. et al. Performance of artificial intelligence-based coronary artery calcium scoring in non-gated chest CT. *Eur. J. Radiol.* **145**, 110034 (2021).
38. Xu, C. et al. Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT. *Quant. Imaging Med. Surg.* **12**, 2684–2695 (2022).
39. Xu, Y. et al. Deep learning-based coronary calcium score derived from non-gated chest CT and major adverse cardiovascular events in patients with type 2 diabetes mellitus. *Cardiovasc. Diabetol.* **24**, 389 (2025).
40. Yin, K. et al. Performance assessment of an artificial intelligence-based coronary artery calcium scoring algorithm in non-gated chest CT scans of different slice thickness. *Quant. Imaging Med. Surg.* **14**, 5708–5720 (2024).
41. Yu, J. et al. Automated total and vessel-specific coronary artery calcium (CAC) quantification on chest CT: direct comparison with CAC scoring on non-contrast cardiac CT. *BMC Med. Imaging* **22**, 177 (2022).
42. Zhang, J., Liu, K., You, C. & Gong, J. Vessel-specific reliability of artificial intelligence-based coronary artery calcium scoring on non-ECG-gated chest CT: a comparative study with ECG-gated cardiac CT. *Clin. Imaging* **125**, 110575 (2025).
43. Anifowose, J. O. et al. Automated opportunistic cardiovascular risk assessment in non-small cell lung cancer patients on routine chest CT using an optimised nnU-net framework. *BMC Med. Imaging* **26**, 179 (2026).
44. Balbi, M. et al. Automated coronary artery calcium and quantitative emphysema in lung cancer screening. *J. Thorac. Imaging* **38**, W52–W63 (2023).
45. Barrett, O. et al. Incremental diagnostic value of artificial intelligence-derived coronary artery calcium in 18F-flurpiridaz positron emission tomography myocardial perfusion imaging. *J. Nucl. Cardiol.* **54**, 102532 (2025).

46. Cano-Espinosa, C., González, G., Washko, G. R., Cazorla, M. & San José Estépar, R. Automated Agatston score computation in non-ECG gated CT scans using deep learning. *Proc. SPIE* **10574**, 105742K (2018).
47. Cano-Espinosa, C. et al. On the relevance of the loss function in the Agatston score regression from non-ECG gated CT scans. in *Image Analysis for Moving Organ, Breast, and Thoracic Images* (eds Stoyanov, D. et al.) 326–334 (Springer, 2018).
48. Cano-Espinosa, C., Gonzalez, G., Washko, G. R., Cazorla, M. & Estepar, R. S. J. Biomarker localization from deep learning regression networks. *IEEE Trans. Med. Imaging* **39**, 2121–2132 (2020).
49. Chamberlin, J. et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. *BMC Med.* **19**, 55 (2021).
50. Ching, K. et al. A manually annotated, open-source dataset for coronary, aortic, and valvular calcium segmentation in non-gated chest CT. *Int. J. Imaging Syst. Technol.* **36**, e70311 (2026).
51. Choi, J. H. et al. Validation of deep learning-based fully automated coronary artery calcium scoring using non-ECG-gated chest CT in patients with cancer. *Front. Oncol.* **12**, 989250 (2022).
52. de Vos, B. D. et al. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans. Med. Imaging* **38**, 2127–2138 (2019).
53. Gernaat, S. A. M. et al. Automatic coronary artery calcium scoring on radiotherapy planning CT scans of breast cancer patients: reproducibility and association with traditional cardiovascular risk factors. *PLoS ONE* **11**, e0167925 (2016).
54. Gernaat, S. A. M. et al. Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of Western and Asian breast cancer patients. *Radiother. Oncol.* **127**, 487–492 (2018).
55. Gonzalez, G., Washko, G. R. & Estepar, R. S. J. Automated Agatston score computation in a large dataset of non ECG-gated chest computed tomography. *IEEE Int. Symp. Biomed. Imaging* 53–57 (2016).
56. Hamelink, I. et al. Validation of artificial intelligence software for automatic calcium scoring in cardiac and chest computed tomography. *Eur. J. Radiol.* **191**, 112323 (2025).
57. Huo, Y. et al. Coronary calcium detection using 3D attention identical dual deep network based on weakly supervised learning. *Proc. SPIE* **10949**, 1094917 (2019).
58. Imuro, S. E. S. et al. Classifying coronary artery calcification risk using radiomics from non-gated CT scans. *IEEE Int. Symp. Biomed. Imaging* 1–4 (2025).
59. Išgum, I. et al. Automatic coronary calcium scoring in low-dose non-ECG-synchronized thoracic CT scans. *Proc. SPIE* **7624**, 76240M (2010).
60. Išgum, I., Prokop, M., Niemeijer, M., Viergever, M. A. & van Ginneken, B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans. Med. Imaging* **31**, 2322–2334 (2012).
61. Išgum, I. et al. Automatic detection of cardiovascular risk in CT attenuation correction maps in Rb-82 PET/CTs. *Proc. SPIE* **9784**, 978405 (2016).
62. Lessmann, N. et al. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. *Proc. SPIE* **9785**, 978511 (2016).
63. Lessmann, N. et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans. Med. Imaging* **37**, 615–625 (2018).
64. Miller, R. J. H. et al. Deep learning coronary artery calcium scores from SPECT/CT attenuation maps improve prediction of major adverse cardiac events. *J. Nucl. Med.* **64**, 652–658 (2023).
65. Mineo, E. et al. Automated coronary artery calcium scoring using deep learning: validation across diverse chest CT protocols. *Acad. Radiol.* **32**, 7177–7186 (2025).
66. Nguyen, A., Liu, J., Mathai, T. S., Grayson, P. C. & Summers, R. M. Automated plaque detection and Agatston score estimation on non-contrast CT scans: a multicenter study. *Proc. SPIE* **12927**, 1292710 (2024).
67. Rathore, S. et al. Fully automated coronary artery calcium score and risk categorization from chest CT using deep learning and multiorgan segmentation: a validation study from the National Lung Screening Trial (NLST). *Int. J. Cardiol. Heart Vasc.* **56**, 101593 (2025).
68. Shadmi, R., Mazo, V., Bregman-Amitai, O. & Elnekave, E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. *IEEE Int. Symp. Biomed. Imaging* 24–28 (2018).
69. Shen, H. et al. Cardiovascular risk stratification by automatic coronary artery calcium scoring on pretreatment chest computed tomography in diffuse large B-cell lymphoma receiving anthracycline-based chemotherapy: a multicenter study. *Circ. Cardiovasc. Imaging* **16**, e014829 (2023).
70. Suh, Y. J. et al. Fully automatic coronary calcium scoring in non-ECG-gated low-dose chest CT: comparison with ECG-gated cardiac CT. *Eur. Radiol.* **33**, 1254–1265 (2023).
71. Takx, R. A. P. et al. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLoS ONE* **9**, e91239 (2014).
72. Tang, H., Moradi, M., Prasanna, P., Wang, H. & Syeda-Mahmood, T. An algorithm for fully automatic detection of calcium in chest CT imaging. *IEEE Int. Symp. Biomed. Imaging* 265–269 (2017).
73. van Velzen, S. G. M. et al. Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology* **295**, 66–79 (2020).
74. van Velzen, S. G. M. et al. Generative models for reproducible coronary calcium scoring. *J. Med. Imaging (Bellingham)* **9**, 052406 (2022).

75. Wang, M. et al. Quantifying the impact of slice thickness on cardiovascular risk stratification in lung cancer screening: a multi-center “RESCUE” study. *Quant. Imaging Med. Surg.* **16**, 568 (2026).
76. Alnasser, T. N. et al. A novel fully automated deep learning model for coronary artery calcification detection on computed tomography. *Diagnostics (Basel)* **16**, 646 (2026).
77. Bonatti, M., Proner, B., Vingiani, V., Valletta, R. & Saba, L. The role of AI assistance in the evaluation of unenhanced chest CT scans in an emergency setting. *Pol. J. Radiol.* **91**, 132–139 (2026).
78. Chamberlin, J. H. et al. A deep convolutional neural network ensemble for composite identification of pulmonary nodules and incidental findings on routine PET/CT. *Clin. Radiol.* **78**, e368–e376 (2023).
79. Kerndt, C. C. et al. Using artificial intelligence to semi-quantitate coronary calcium as an ‘incidentaloma’ on non-gated, non-contrast CT scans, a single-center descriptive study in West Michigan. *Spartan Med. Res. J.* **8**, 89132 (2023).
80. Park, C. et al. Coronary artery calcification on low-dose lung cancer screening CT in South Korea: visual and artificial intelligence–based assessment and association with cardiovascular events. *AJR Am. J. Roentgenol.* **222**, e2430852 (2024).
81. Shear, B. et al. Assessing the diagnostic accuracy and prognostic utility of artificial intelligence detection and grading of coronary artery calcification on nongated computed tomography (CT) thorax. *Clin. Radiol.* **85**, 106909 (2025).
82. Smith, C. et al. Application of an artificial intelligence ensemble for detection of important secondary findings on lung ventilation and perfusion SPECT-CT. *Clin. Imaging* **100**, 24–29 (2023).
83. Wu, K.-C. et al. Application of deep learning in automated localization and interpretation of coronary artery calcification in oncological PET/CT scans. *Int. J. Cardiovasc. Imaging* **41**, 453–465 (2025).
84. Xie, Y. et al. Coronary artery calcification identification and labeling in low-dose chest CT images. *Proc. SPIE* **10134**, 101340L (2017).
85. Atkins, K. M. et al. Elevated coronary artery calcium quantified by a validated deep learning model from lung cancer radiotherapy planning scans predicts mortality. *JCO Clin. Cancer Inform.* **6**, e2100095 (2022).
86. Barda, N. et al. Improving cardiovascular disease prediction using automated coronary artery calcium scoring from existing chest CTs. *J. Digit. Imaging* **35**, 962–969 (2022).
87. de Vos, B. D. et al. Automatic machine learning based prediction of cardiovascular events in lung cancer screening data. *Proc. SPIE* **9414**, 94140D (2015).
88. de Vos, B. D., Lessmann, N., de Jong, P. A. & Išgum, I. Deep learning–quantified calcium scores for automatic cardiovascular mortality prediction at lung screening low-dose CT. *Radiol. Cardiothorac. Imaging* **3**, e190219 (2021).
89. Dekker, M. et al. The prognostic value of automated coronary calcium derived by a deep learning approach on non-ECG gated CT images from 82Rb-PET/CT myocardial perfusion imaging. *Int. J. Cardiol.* **329**, 9–15 (2021).
90. Kim, E.-J., Cho, S. W., Yang, J.-H. & Jeong, W. G. Artificial intelligence-assisted longitudinal assessment of coronary artery calcification in the Korean lung cancer screening CT program. *Clin. Imaging* **125**, 110558 (2025).
91. Kim, K. et al. Automated coronary artery calcium scoring in patients with breast cancer to assess the risk of heart disease following adjuvant radiation therapy. *Breast* **65**, 77–83 (2022).
92. Kim, N. Y., Kim, Y.-H., Lee, J. E. & Suh, Y. J. Opportunistic assessment of coronary artery calcium volume and density from non-electrocardiogram-gated chest CT using artificial intelligence: prognostic implications in a screening cohort. *Korean J. Radiol.* **27**, 532–542 (2026).
93. Kim, S. et al. Association between automated coronary artery calcium from routine chest computed tomography scans and cardiovascular risk in patients with colorectal or gastric cancer. *Circ. Cardiovasc. Qual. Outcomes* **18**, e011656 (2025).
94. Lee, J. E. et al. Long-term prognostic implications of thoracic aortic calcification on CT using artificial intelligence–based quantification in a screening population: a two-center study. *AJR Am. J. Roentgenol.* **224**, e2532697 (2025).
95. Lessmann, N. et al. Sex differences in coronary artery and thoracic aorta calcification and their association with cardiovascular mortality in heavy smokers. *JACC Cardiovasc. Imaging* **12**, 1808–1817 (2019).
96. Mets, O. M. et al. Lung cancer screening CT-based prediction of cardiovascular events. *JACC Cardiovasc. Imaging* **6**, 899–907 (2013).
97. Sato, H. et al. The impact of the coronary artery calcium score on the clinical outcomes in patients with acute myocardial infarction. *J. Clin. Med.* **13**, 7136 (2024).
98. Stemmer, A. et al. Using machine learning algorithms to review computed tomography scans and assess risk for cardiovascular disease: retrospective analysis from the National Lung Screening Trial (NLST). *PLoS ONE* **15**, e0236021 (2020).
99. Takx, R. A. P. et al. Quantification of coronary artery calcium in nongated CT to predict cardiovascular events in male lung cancer screening participants: results of the NELSON study. *J. Cardiovasc. Comput. Tomogr.* **9**, 50–57 (2015).
100. Takx, R. A. P. et al. Pulmonary function and CT biomarkers as risk factors for cardiovascular events in male lung cancer screening participants: the NELSON study. *Eur. Radiol.* **25**, 65–71 (2015).
101. Weber, B. N. et al. Prevalence and prognostic value of incidentally detected coronary artery calcium using artificial intelligence among individuals with immune-mediated inflammatory diseases. *JACC Cardiovasc. Imaging* **19**, 64–75 (2026).

102. Williams, M. C. et al. Automated vessel-specific coronary artery calcification quantification with deep learning in a large multi-centre registry. *Eur. Heart J. Cardiovasc. Imaging* **25**, 976–985 (2024).
103. Xing, Z. et al. Incremental value of non-gated chest CT coronary artery calcium score in predicting major adverse cardiovascular events by GRACE score after percutaneous coronary intervention in patients with acute coronary syndrome. *Clin. Cardiol.* **49**, e70242 (2026).
104. Yang, K. et al. Association between deep learning-based coronary artery calcium score on non-gated chest CT and progression of chronic kidney disease: a retrospective observational cohort study. *BMC Med. Imaging* **26**, 291 (2026).
105. Zhou, J. et al. Hidden risk in normal myocardial perfusion scans: AI-detected proximal coronary calcium on CT attenuation maps improves prognosis. *Eur. J. Nucl. Med. Mol. Imaging* https://doi.org/10.1007/s00259-026-08084-x (2026).
106. Bouladian, S. G. et al. Statin therapy persistence following opportunistic screening for coronary artery calcium on nongated chest CTs (NOTIFY-EXTEND). *JACC Cardiovasc. Imaging* **18**, 841–843 (2025).
107. Chernina, V. Y. et al. Analysis of the diagnostic and economic impact of the combined artificial intelligence algorithm for analysis of 10 pathological findings on chest computed tomography. *Digit. Diagn.* **4**, 105–132 (2023).
108. Bunk, S. et al. Repeatability of AI-quantified incidental findings on lung cancer screening CT scans in the NELSON trial. *Eur. Radiol.* https://doi.org/10.1007/s00330-026-12669-3 (2026).
109. Hamelink, I., van Tuinen, M., Kwee, T. C., van Ooijen, P. M. A. & Vliegenthart, R. Repeatability of AI-based, automatic measurement of vertebral and cardiovascular imaging biomarkers in low-dose chest CT: the ImaLife cohort. *Eur. Radiol.* **35**, 3833–3841 (2025).
110. Ledda, R. E. et al. Coronary calcium score and emphysema extent on different CT radiation dose protocols in lung cancer screening. *Eur. Radiol.* **35**, 3781–3787 (2025).
111. Yacoub, B. et al. Performance of an artificial intelligence-based platform against clinical radiology reports for the evaluation of noncontrast chest CT. *Acad. Radiol.* **29**, S108–S117 (2022).
112. Liu, Y. et al. Validation of artificial intelligence-based non-gated chest CT for coronary artery calcium scoring across multiple CT scanners. *J. Imaging Inform. Med.* https://doi.org/10.1007/s10278-026-02288-1 (2026).
113. Shanbhag, A. et al. Cross-domain knowledge transfer from expert-annotated gated CT via synthetic ungated CT improves coronary artery calcium scoring on CT attenuation correction scans. Preprint at https://doi.org/10.64898/2026.07.02.26356002 (2026).
114. Ochs, K. et al. Agreement and workflow efficiency of AI-based coronary artery calcification quantification in lung cancer screening: Comparison with semi-automated and visual assessment. *Eur. J. Radiol. Open* **17**, 100799 (2026).
115. Noureldine, H. A. et al. Assessing the prognostic value of artificial intelligence for opportunistic prediction of coronary artery calcium score on routine non-gated computed tomography in patients treated with immune checkpoint inhibitors. *Digit. Health* **12**, 20552076261480207 (2026).
116. Liu, Y. K. et al. Artificial intelligence-based chest computed tomography assessment of bone, vascular, and epicardial adipose changes in type 2 diabetes. *World J. Diabetes* **17**, 121422 (2026).
117. Yi, J. et al. Beyond BMI: an interpretable integrated body composition index from low-dose chest CT for all-cause mortality risk stratification: a multicentre study. Preprint at https://doi.org/10.64898/2026.08.05.26359437 (2026).
118. Cheng, D. et al. Deep learning-derived cardiac phenotyping from routine non-gated chest CT for mortality prediction and risk stratification in chronic kidney disease: a development and external validation cohort study. Preprint at https://doi.org/10.2139/ssrn.7040888 (2026).
119. Hussain, M. et al. Deep learning for automated coronary artery calcium scoring in chest CTs. *Int. J. Drug Deliv. Technol.* **16**, 1535–1541 (2026).
120. Peng, A. W. et al. Association of coronary artery calcium detected by routine ungated CT imaging with cardiovascular outcomes. *J. Am. Coll. Cardiol.* **82**, 1192–1202 (2023).
121. Parsa, S. et al. Measurement and application of incidentally detected coronary calcium: JACC review topic of the week. *J. Am. Coll. Cardiol.* **83**, 1557–1567 (2024).
122. Nasir, K. & Blankstein, R. Transforming the cardiovascular disease prevention paradigm: see disease, treat disease. *JAMA* **333**, 1398–1400 (2025).
123. Muhlestein, J. B. et al. Coronary artery calcium versus pooled cohort equations score for primary prevention guidance: randomized feasibility trial. *JACC Cardiovasc. Imaging* **15**, 843–855 (2022).
124. Whitmore, K. et al. Impact of patient visualization of cardiovascular images on modification of cardiovascular risk factors: systematic review and meta-analysis. *JACC Cardiovasc. Imaging* **16**, 1069–1081 (2023).
125. Verghese, D. et al. Predictive value of coronary artery calcium score from nongated chest CT scans compared with gated scans and incidence of cardiovascular events. *Circulation* **154**, 302–312 (2026).
126. Nerlekar, N. et al. Effects of combining coronary calcium score with treatment on plaque progression in familial coronary artery disease: a randomized clinical trial. *JAMA* **333**, 1403–1412 (2025).
127. Williams, M. C. et al. Reporting incidental coronary, aortic valve and cardiac calcification on non-gated thoracic computed tomography, a consensus statement from the BSCI/BSCCT and BSTI. *Br. J. Radiol.* **94**, 20200894 (2021).
128. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).

129. McInnes, M. D. F. et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* **319**, 388–396 (2018).
130. Rethlefsen, M. L. et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst. Rev.* **10**, 39 (2021).
131. Whiting, P. F. et al. QUADAS-3: a revised tool for the quality assessment of diagnostic test accuracy studies. *Ann. Intern. Med.* **179**, 548–555 (2026).
132. Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P. & Bombardier, C. Assessing bias in studies of prognostic factors. *Ann. Intern. Med.* **158**, 280–286 (2013).
133. Sterne, J. A. C. et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* **366**, l4898 (2019).
134. Schünemann, H. J. et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J. Clin. Epidemiol.* **122**, 129–141 (2020).
135. Schünemann, H. J. et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J. Clin. Epidemiol.* **122**, 142–152 (2020).
136. Mustafa, R. A. et al. GRADE guidance 37: rating imprecision in a body of evidence on test accuracy. *J. Clin. Epidemiol.* **165**, 111189 (2024).
137. Foroutan, F. et al. GRADE guidelines 28: use of GRADE for the assessment of evidence about prognostic factors. *J. Clin. Epidemiol.* **121**, 62–70 (2020).
138. Murad, M. H. et al. GRADE guidance 38: updated guidance for rating up certainty of evidence due to a dose–response gradient. *J. Clin. Epidemiol.* **164**, 45–53 (2023).
139. Guyatt, G. H. et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J. Clin. Epidemiol.* **64**, 383–394 (2011).
140. Schünemann, H. J. et al. Completing 'Summary of findings' tables and grading the certainty of the evidence. in *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5.1 (eds Higgins, J. P. T. et al.) Ch. 14 (Cochrane, 2024).
141. Reitsma, J. B. et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **58**, 982–990 (2005).
142. Rutter, C. M. & Gatsonis, C. A. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat. Med.* **20**, 2865–2884 (2001).
143. Deeks, J. J., Macaskill, P. & Irwig, L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J. Clin. Epidemiol.* **58**, 882–893 (2005).
144. Bipat, S., Zwinderman, A. H., Bossuyt, P. M. M. & Stoker, J. Multivariate random-effects approach: for meta-analysis of cancer staging studies. *Acad. Radiol.* **14**, 974–984 (2007).
145. Hamza, T. H., Arends, L. R., van Houwelingen, H. C. & Stijnen, T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med. Res. Methodol.* **9**, 73 (2009).
146. Met, R., Bipat, S., Legemate, D. A., Reekers, J. A. & Koelemay, M. J. W. Diagnostic performance of computed tomography angiography in peripheral arterial disease: a systematic review and meta-analysis. *JAMA* **301**, 415–424 (2009).
147. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
148. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
149. IntHout, J., Ioannidis, J. P. A. & Borm, G. F. The Hartung–Knapp–Sidik–Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian–Laird method. *BMC Med. Res. Methodol.* **14**, 25 (2014).
150. Knapp, G. & Hartung, J. Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693–2710 (2003).
151. R Core Team. *R: A Language and Environment for Statistical Computing*, version 4.5.3 (R Foundation for Statistical Computing, 2026).
152. Doebler, P. mada: meta-analysis of diagnostic accuracy. R package version 0.5.12 https://CRAN.R-project.org/package=mada (2025).
153. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
154. Gabry, J., Češnovar, R. & Johnson, A. cmdstanr: R interface to CmdStan. R package version 0.9.0 https://mc-stan.org/cmdstanr/ (2025).

781 **Table 1. Reports contributing to at least one meta-analysis (n = 59)**

Report	Region	Setting	Participants	Non-contrast CT confirmed	AI system	Validation	Syntheses
Dobrolinska et al, 2023[15]	Europe	PET/CT attenuation-correction LDCT; suspected CAD	213	Yes	syngo.via CT CaScoring (commercial)	External	G, K
Dobrolinska et al, 2024[16]	Europe	PET/CT attenuation-correction LDCT; suspected CAD	572; 507	Yes	Academic deep-learning model	Internal; prognostic	G, P, K
Eng et al, 2021[17]	Multinational	Routine chest CT; multicentre validation cohorts	391	Yes	Academic deep-learning model	External and internal	G, K
Gokmen et al, 2026[18]	North America	Non-gated chest CT with paired gated cohorts	205	No	Academic: CARD-ViT	External	G
Hagopian et al, 2025[19]	North America	Routine non-contrast chest CT; Veterans Affairs	795	Yes	Academic: AI-CAC (open source)	Internal; held-out test set	G, K
Kang et al, 2023[22]	East Asia	Low-dose chest CT with paired gated CT	567; 453	No	AVIEW CAC (commercial)	External	G, K
Kim et al, 2025a[23]	East Asia	Low-dose chest CT with paired gated CT	225; 93	Yes	AVIEW CAC with image conversion (commercial)	Held-out test set	G, K
Kim et al, 2025b[24]	Multinational	Low-dose chest CT with paired gated CT	652; 199	No	ClariCardio (commercial)	External	G, M, K
Li et al, 2023[25]	North America	Public chest CT dataset (COCA)	659; 63; 89	No	Academic deep-learning model	Internal	K
Ludwig et al, 2026[27]	North America	Mixed cohorts: calcium scoring, lung screening, transplant candidates	495; 245; 125	Yes	DL-CAC (commercial)	External	G
McKinney et al, 2025[29]	North America	Routine non-gated chest CT; five hospitals	491; 331	Yes	ClearRead CT CAC (commercial)	External	G, M, K
Morf et al, 2022[30]	Europe	Oncology FDG PET/CT attenuation-correction CT	100	Yes	AVIEW CAC (commercial)	External	G, K
Oh et al, 2024[31]	East Asia	PET/CT attenuation-correction CT with paired gated CT	210	Yes	AVIEW CAC (commercial)	External	G, K
Pieszko et al, 2022[32]	North America	PET myocardial perfusion CT attenuation maps	507	No	Academic deep-learning model	External	G, K
Pieszko et al, 2023[33]	North America	PET myocardial perfusion CT attenuation maps	4,331; 2,737; 2,502	No	Academic deep-learning model	External; prognostic	G, P, K
van Assen et al, 2021[35]	North America	Chest CT with paired cardiac CT	93	Yes	AI-Rad Companion Chest CT (commercial)	External	G, K
Xu et al, 2021[37]	East Asia	Non-gated chest CT with paired gated CT	901	Yes	CACScoreDoc (commercial)	Held-out test set	G, K
Xu et al, 2022[38]	East Asia	Routine chest CT with paired gated CT	144; 117	Yes	Academic deep-learning model	External and internal	G, K

Report	Region	Setting	Participants	Non-contrast CT confirmed	AI system	Validation	Syntheses
Xu et al, 2025[39]	East Asia	Inpatients with type 2 diabetes; chest CT	1,041; 694; 506; 400	No	uAI Discover (commercial)	External	G, K
Yin et al, 2024[40]	East Asia	Suspected CAD; low-dose chest CT with paired gated CT	112	Yes	CACScoreDoc (commercial)	Unclear	G, K
Yu et al, 2022[41]	East Asia	Known or suspected CAD; chest CT with paired gated CT	405	Yes	CACScoreDoc (commercial)	External	G, K
Anifowose et al, 2026[43]	Multinational	Non-small-cell lung cancer; chest CT and PET-CT	97; 95; 81; 50; 226	Yes	Academic: OncCalc	External	M
Balbi et al, 2023[44]	Europe	Lung-cancer screening LDCT (BioMILD)	4,098; 100	No	AVIEW (commercial)	External	M
Barrett et al, 2025[45]	Multinational	PET myocardial perfusion CT attenuation maps	498	Yes	Academic deep-learning model	External	M, K
Cano-Espinosa et al, 2018a[46]	North America	Smokers; chest CT (COPDGene)	1,000	Yes	Academic deep-learning model	Internal	M
Cano-Espinosa et al, 2020[48]	North America	Smokers; chest CT (COPDGene)	988	Yes	Academic deep-learning model	Internal	M
Chamberlin et al, 2021[49]	North America	Lung-cancer screening LDCT	117; 96	Yes	AI-Rad Companion (commercial)	External	M
Choi et al, 2022[51]	East Asia	Oncology; routine non-gated chest CT	913	Yes	AVIEW CAC (commercial)	External	M, K
de Vos et al, 2019[52]	Multinational	Chest CT and cardiac CT datasets	506	Yes	Academic deep-learning model	Internal	M, K
Gernaat et al, 2016[53]	Europe	Breast cancer radiotherapy-planning CT	163; 294	Yes	Academic machine-learning algorithm	External	M, K
González et al, 2016[55]	North America	Smokers; chest CT (COPDGene)	1,000	No	Academic machine-learning algorithm	Internal	M
Hamelink et al, 2025b[56]	Europe	General population; low-dose chest CT (ImaLife)	684	Yes	AI-Rad Companion prototype (commercial)	External	M, K
Huo et al, 2019[57]	North America	Non-contrast chest CT; hospital archive	206; 5,075	Yes	Academic: AID-Net	Internal	M
Išgum et al, 2010[59]	Europe	Lung-cancer screening LDCT (NELSON)	121; 35	Yes	Academic machine-learning algorithm	Internal	M, K
Išgum et al, 2012[60]	Europe	Lung-cancer screening LDCT (NELSON)	231	Yes	Academic machine-learning algorithm	Internal	M
Išgum et al, 2016[61]	North America	Rb-82 PET/CT attenuation-correction CT	33	No	Academic machine-learning algorithm	Internal	M
Lessmann et al, 2016[62]	Europe	Lung-cancer screening LDCT	1,028; 231	Yes	Academic deep-learning model	Internal; held-out test set	M, K
Lessmann et al, 2018[63]	North America	Lung-cancer screening LDCT	1,459; 310	Yes	Academic deep-learning model	Internal	M
Miller et al, 2023[64]	North America	SPECT/CT myocardial perfusion attenuation maps	2,271	Yes	Academic deep-learning model	External; prognostic	M, P, K

Report	Region	Setting	Participants	Non-contrast CT confirmed	AI system	Validation	Syntheses
Mineo et al, 2025[65]	Multinational	Routine non-cardiac chest CT; multiple protocols	111; 212	No	Academic deep-learning model	External and internal	M, K
Rathore et al, 2025[67]	North America	Lung-cancer screening LDCT (NLST)	1,442	Yes	KALCIUM / KardioNet (commercial)	Internal; held-out test set	M, K
Shadmi et al, 2018[68]	Multinational	Non-contrast chest CT; hospital cohort	203; 14,365	Yes	FC-DenseNet (commercial)	Held-out test set	M, K
Suh et al, 2023[70]	East Asia	Low-dose chest CT with paired gated CT	452	No	AVIEW CAC (commercial)	External	M, K
Takx et al, 2014[71]	Europe	Lung-cancer screening LDCT	1,749	Yes	Academic machine-learning algorithm	Internal	M, K
van Velzen et al, 2020[73]	Multinational	Multiple CT datasets incl. lung screening and PET/CT	7,240	Yes	Academic deep-learning model	External and internal	M
Wang et al, 2026[75]	Multinational	Multiple datasets incl. COCA and routine chest CT	2,914	No	Academic deep-learning model	External	K
Zeleznik et al, 2021[7]	North America	Lung screening and trial cohorts (NLST, FHS, PROMISE)	396	Yes	Academic deep-learning model	External	M
Alnasser et al, 2026[76]	Europe	Pulmonary hypertension registry; non-gated CT	473	Yes	Academic deep-learning model	External	V
Bonatti et al, 2026[77]	Europe	Emergency department unenhanced chest CT	90	Yes	xAID Chest (commercial)	Held-out test set	V
Chamberlin et al, 2023[78]	North America	Oncology PET/CT	189	Yes	AI-Rad Companion (commercial)	External	V
Shear et al, 2025[81]	Europe	Routine non-gated chest CT	291; 234	Yes	4DMedical CAC (commercial)	External	V
Lee et al, 2025[94]	East Asia	Asymptomatic health screening; routine chest CT	5,342	Yes	ClariCardio (commercial)	External; prognostic	P
Williams et al, 2024[102]	North America	Attenuation-correction CT; multicentre registry	5,969	Yes	Academic deep-learning model	External; prognostic	P
Yang et al, 2026[104]	East Asia	Chronic kidney disease; routine chest CT	509	No	Academic: DL-CACS	Prognostic	P
Dudum et al, 2026[12]	North America	Randomised trial of AI-CAC notification (NOTIFY-PICTURE)	202	Yes	FDA-cleared AI-CAC (commercial)	Randomised trial	I
Sandhu et al, 2023[11]	North America	Randomised trial of AI-CAC notification (NOTIFY-1)	173	Yes	Academic deep-learning model	Randomised trial	I
Liu Y et al, 2026[112]	East Asia	Routine chest CT; coronary or pulmonary disease	1,334	Yes	United Imaging software (commercial)	Unclear	G
Ochs et al, 2026[114]	Europe	Lung-cancer screening LDCT	323	Yes	ClearRead CT CAC (commercial)	External	M, K
Hussain et al, 2026[119]	South Asia	Clinically indicated non-contrast chest CT	210	Yes	NR	Unclear	M

Syntheses: G, paired ECG-gated reference; M, same-scan manual reference; V, expert visual reference; K, weighted-κ category agreement; P, prognostic MACE synthesis; I, randomised implementation synthesis.
Reports could contribute to more than one synthesis: 21 contributed to the paired ECG-gated synthesis, 29 to the same-scan manual synthesis, 4 to the expert visual synthesis, 34 to the weighted-κ synthesis, 6 to the

MACE synthesis, and 2 to the implementation synthesis. Participants are analysis-cohort sizes; several values indicate separate cohorts. Non-contrast CT confirmed indicates that the analysed non-gated examinations were documented as non-contrast. AI, artificial intelligence; CAC, coronary artery calcium; CAD, coronary artery disease; CT, computed tomography; ECG, electrocardiogram; LDCT, low-dose CT; MACE, major adverse cardiovascular events; NR, not reported; PET, positron emission tomography; SPECT, single-photon emission CT.

Table 2. Pooled diagnostic performance of AI-derived CAC on non-gated chest CT by reference standard and threshold

Reference standard	Threshold	Reports	Estimates	Participants	Sensitivity, % (95% CI)	Sensitivity 95% PI	Specificity, % (95% CI)	Specificity 95% PI	I², % (Se / Sp)	LR+ (95% CI)	LR− (95% CI)	GRADE certainty (Se / Sp)
Paired ECG-gated CT	CAC >0	17	24	9,393	87.1 (83.2–90.2)	62.6–96.4	92.3 (87.2–95.4)	50.2–99.3	93 / 94	11.2 (6.8–18.9)	0.14 (0.11–0.18)	Low / Very low
Paired ECG-gated CT	CAC ≥100	19	26	10,940	83.1 (78.5–86.8)	56.4–94.9	96.3 (94.2–97.7)	76.1–99.5	91 / 92	22.4 (14.4–34.8)	0.18 (0.14–0.22)	Low / Low
Paired ECG-gated CT	CAC ≥300	4	6	1,727	80.5 (61.5–91.5)	29.3–97.6	98.9 (96.2–99.7)	81.0–99.9	83 / 79	74.2 (23.6–217.9)	0.20 (0.09–0.39)	Very low / Low
Paired ECG-gated CT	CAC ≥400	17	22	9,543	80.1 (74.3–84.9)	52.7–93.6	97.7 (96.2–98.7)	84.3–99.7	85 / 90	35.4 (21.1–59.2)	0.20 (0.16–0.26)	— / —
Same-scan manual scoring	CAC >0	20	22	10,878	93.8 (91.0–95.7)	73.3–98.8	88.7 (83.4–92.4)	53.5–98.2	93 / 94	—	—	— / —
Same-scan manual scoring	CAC ≥100	24	25	14,371	93.3 (91.1–95.0)	79.2–98.1	97.5 (96.0–98.4)	83.2–99.7	85 / 91	—	—	— / —
Same-scan manual scoring	CAC ≥300	5	6	1,925	90.5 (82.0–95.2)	69.3–97.6	98.5 (97.1–99.3)	94.6–99.6	56 / 62	—	—	— / —
Same-scan manual scoring	CAC ≥400	20	20	12,742	90.3 (86.7–93.0)	72.1–97.1	98.7 (97.7–99.3)	88.4–99.9	80 / 91	—	—	— / —
Expert visual assessment	CAC >0	4	4	1,043	95.8 (93.9–97.1)	93.8–97.2	90.3 (68.0–97.6)	27.9–99.6	0 / 93	—	—	— / —

Pooled estimates are from bivariate random-effects models fitted separately for each reference standard and threshold; thresholds are in Agatston units. Likelihood ratios were derived from the joint model covariance and GRADE certainty was assessed only for the paired ECG-gated reference at CAC >0, ≥100, and ≥300 (Supplementary Table 7); cells marked — were not calculated. I² is the univariate between-study heterogeneity for sensitivity (Se) and specificity (Sp). CI, confidence interval; LR+, positive likelihood ratio; LR−, negative likelihood ratio; PI, prediction interval.

Table 3. AI-enabled opportunistic CAC pathways, preventive treatment, and downstream care

Study	Design and population	AI-to-action pathway	Preventive-treatment effect	Downstream care and clinical outcomes
NOTIFY-1, Sandhu et al, 2023[11]	Single-centre randomised quality-improvement trial; n = 173; statin-naive adults without known ASCVD	Retrospective AI-CAC screening → electronic-record exclusions and chart review → radiologist confirmation → clinician notification → patient notification with CAC image and statin guidance	Statin prescription: 44/86 (51.2%) vs 6/87 (6.9%); RR 7.42 (95% CI 3.34–16.50)	More stress testing, new cardiology encounters, and CAD testing; no cardiovascular-event endpoint
NOTIFY-PICTURE, Dudum et al, 2026[12]	Single-centre pragmatic prospective randomised trial; n = 202; lipid-lowering-naive adults	Near-real-time FDA-cleared AI-CAC screening → eligibility review → clinician and patient notification with CAC image and guideline-based treatment guidance; no study-level confirmation step reported	Statin prescription: 44/100 (44.0%) vs 10/102 (9.8%); RR 4.49 (2.39–8.42)	No increase in health-care utilisation reported; no cardiovascular-event endpoint
NOTIFY-EXTEND, Bouladian et al, 2025[106]	Linked single-health-system follow-up of NOTIFY-1; n = 163; no concurrent control group	Longitudinal follow-up after AI detection, clinician/patient notification, and preventive-treatment review	Statin prescribed in 75/163 (46%) by 6 months and 86/163 (53%) by 12 months; 87% of initiators (75 of 86) were still on treatment at 18 months	Testing and clinical visits increased after notification; no cardiovascular-event endpoint

The two randomised studies evaluated bundled AI-enabled care pathways rather than the isolated effect of the AI algorithm. A secondary fixed-effect synthesis of six-month statin prescribing yielded RR 5.44 (95% CI 3.32–8.92). High GRADE certainty applies to six-month statin prescription in the two studied pathways, not to cardiovascular outcomes or an isolated AI effect. ASCVD, atherosclerotic cardiovascular disease; CAC, coronary artery calcium; CAD, coronary artery disease; CI, confidence interval; RR, risk ratio.

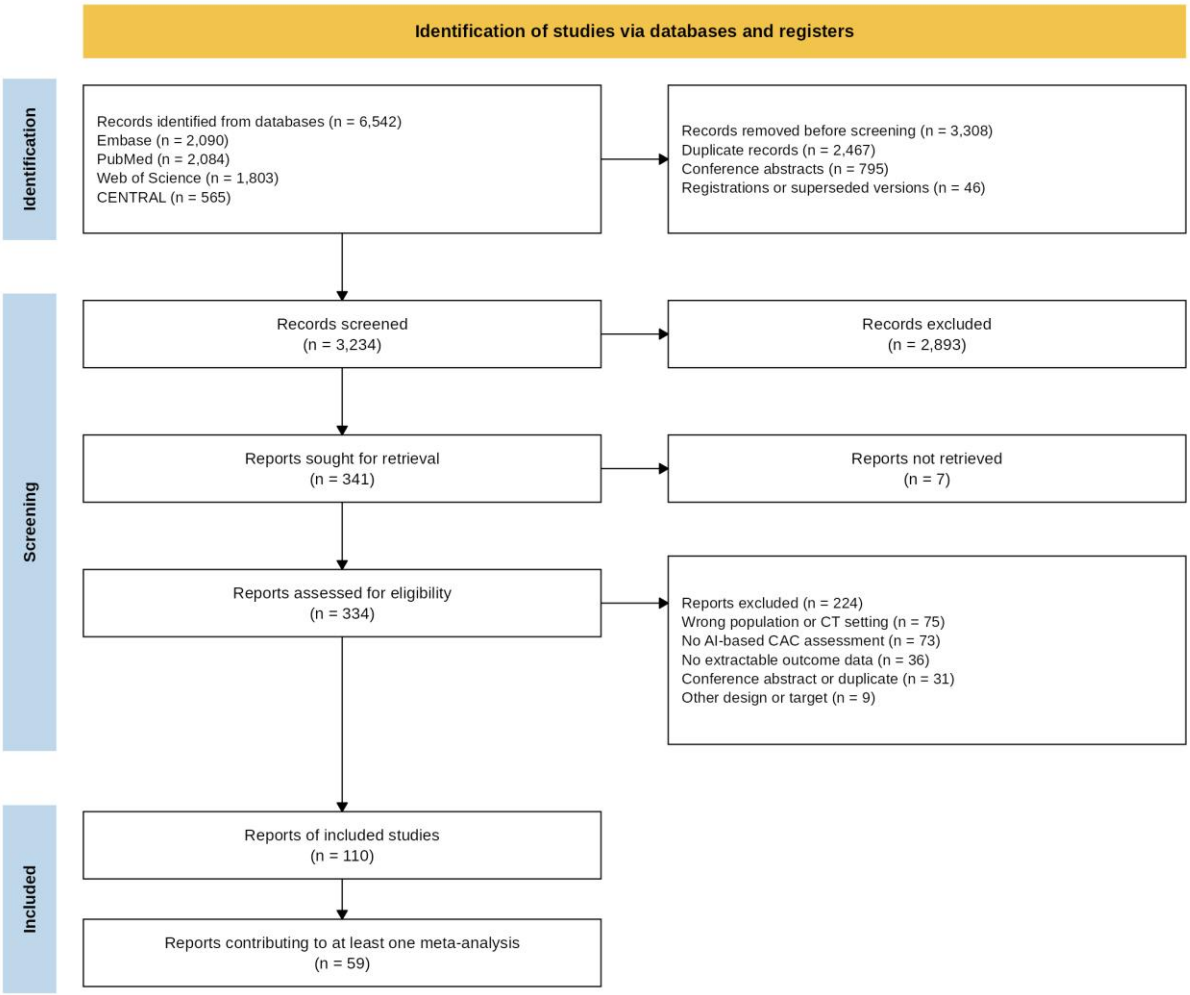


Figure 1. Study selection. PRISMA 2020 flow diagram for searches to 2 October 2026. In total, 110 reports were included and 59 contributed to at least one meta-analysis. Related reports and overlapping cohorts were linked within each analysis to prevent double counting.



Figure 2. Risk of bias of the selected estimates. Bars show the distribution of final judgments across eligible estimates within each evidence domain. **a**, Paired ECG-gated reference (QUADAS-3; 78 estimates). **b**, Same-scan manual reference (QUADAS-3; 73 estimates). **c**, Expert visual reference (QUADAS-3; four estimates). **d**, Prognostic evidence (QUIPS; 16 MACE estimates). **e**, Randomised implementation trials (RoB 2; two trials). QUADAS-3 judgments were Low, High, or Insufficient information; QUIPS judgments were Low, Moderate, or High; RoB 2 judgments were Low, Some concerns, or High. Applicability judgments are shown in Supplementary Fig. 1. MACE, major adverse cardiovascular events.

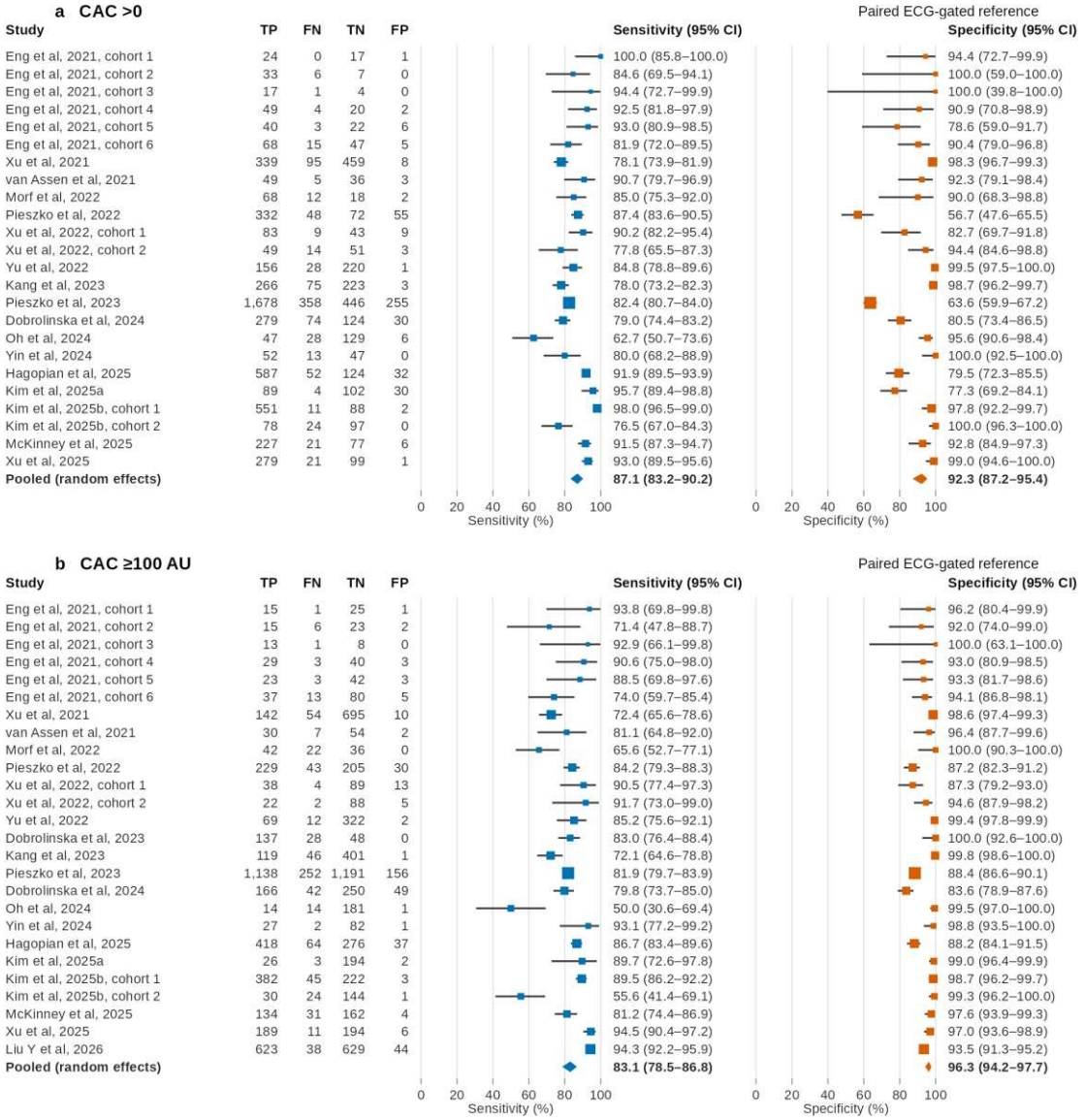


Figure 3. Sensitivity and specificity of AI-derived CAC against paired ECG-gated CT at CAC >0 and ≥100 AU. **a**, CAC >0 (24 estimates from 17 reports; 9,393 participants). **b**, CAC ≥100 AU (26 estimates from 19 reports; 10,940 participants). Squares show estimate-level sensitivity and specificity with exact binomial 95% confidence intervals and are scaled by the square root of the number of participants; diamonds show pooled estimates from bivariate random-effects models with 95% confidence intervals. Columns give true positives, false negatives, true negatives, and false positives. Cohort numbers distinguish independent estimates from the same report. AI, artificial intelligence; AU, Agatston units; CAC, coronary artery calcium; CT, computed tomography; ECG, electrocardiogram.

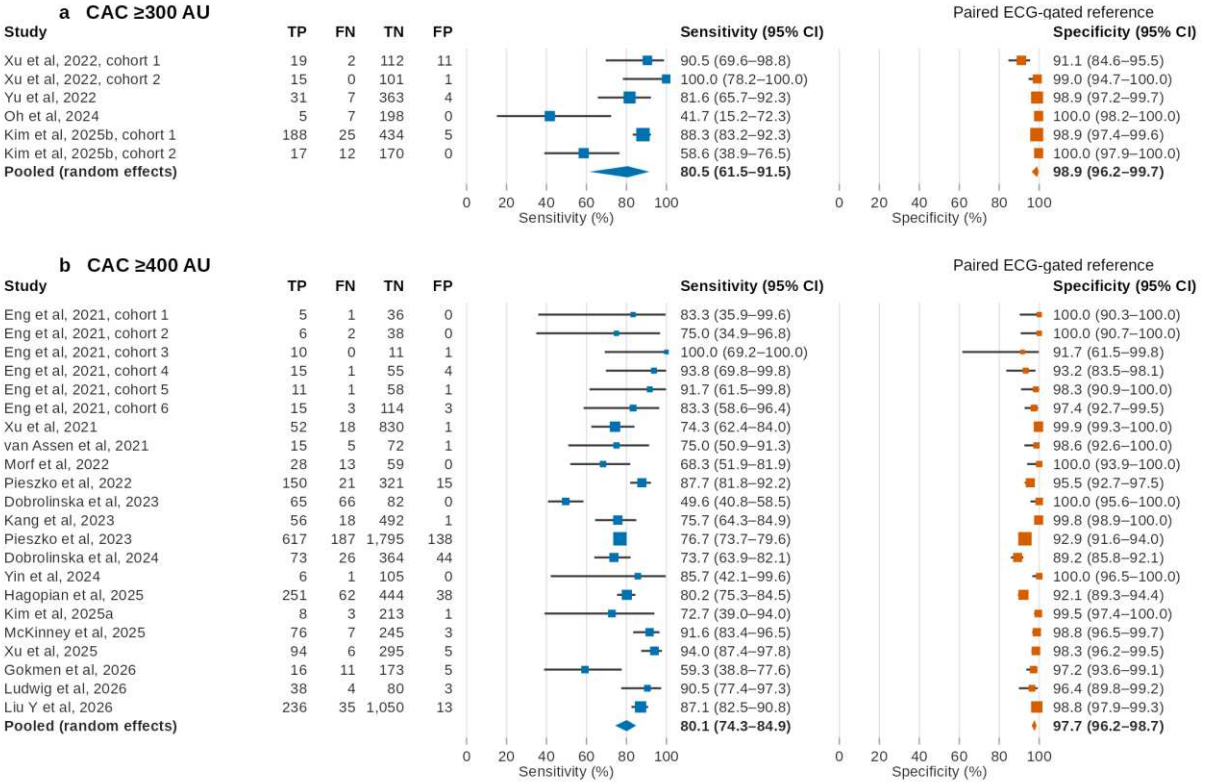


Figure 4. Sensitivity and specificity of AI-derived CAC against paired ECG-gated CT at CAC ≥300 and ≥400 AU. **a**, CAC ≥300 AU (six estimates from four reports; 1,727 participants). **b**, CAC ≥400 AU (22 estimates from 17 reports; 9,543 participants). Plotting conventions as in Fig. 3.

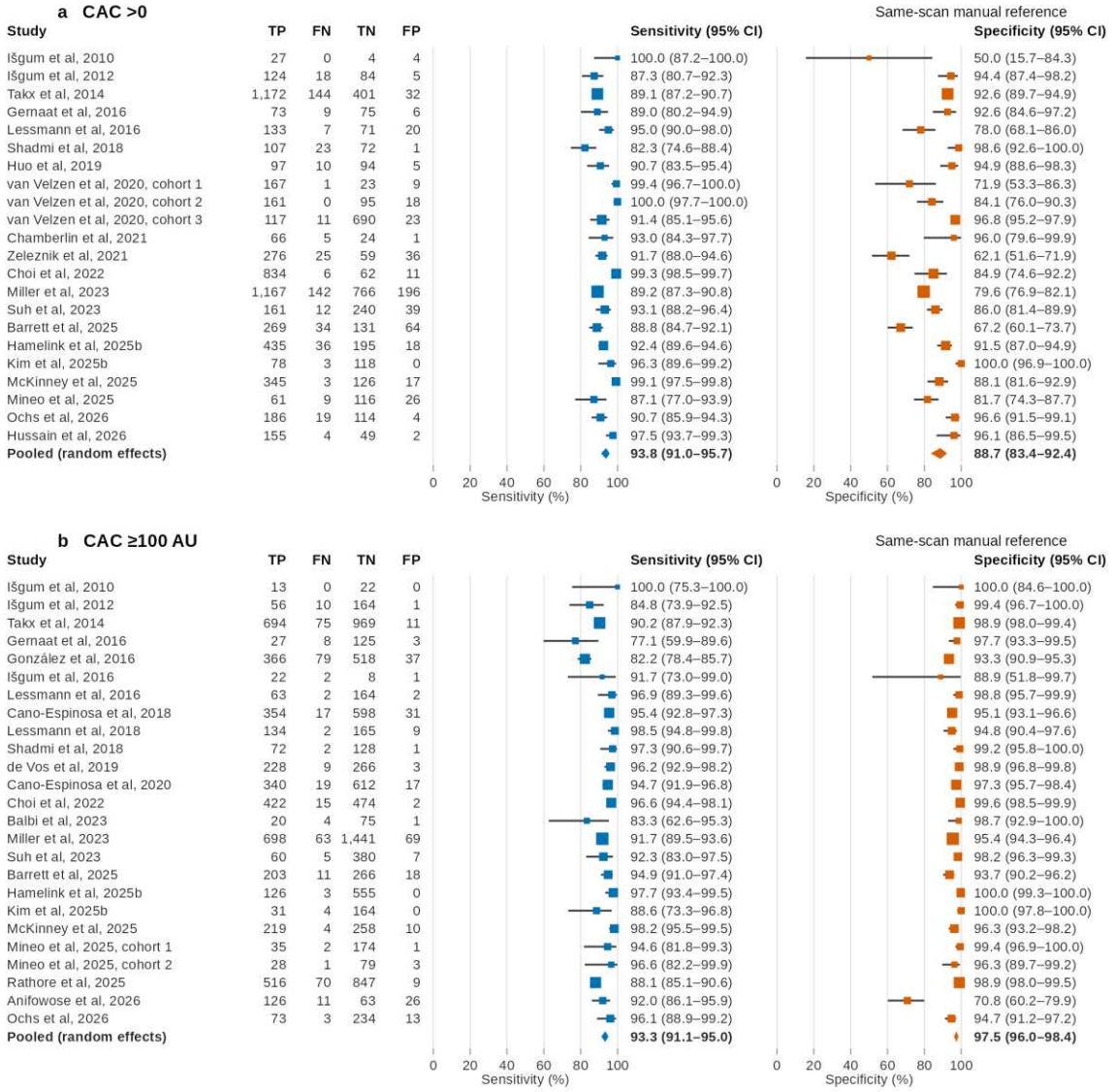


Figure 5. Sensitivity and specificity of AI-derived CAC against manual scoring of the same non-gated CT at CAC >0 and ≥100 AU. a, CAC >0 (22 estimates from 20 reports; 10,878 participants). **b**, CAC ≥100 AU (25 estimates from 24 reports; 14,371 participants). Plotting conventions as in Fig. 3.

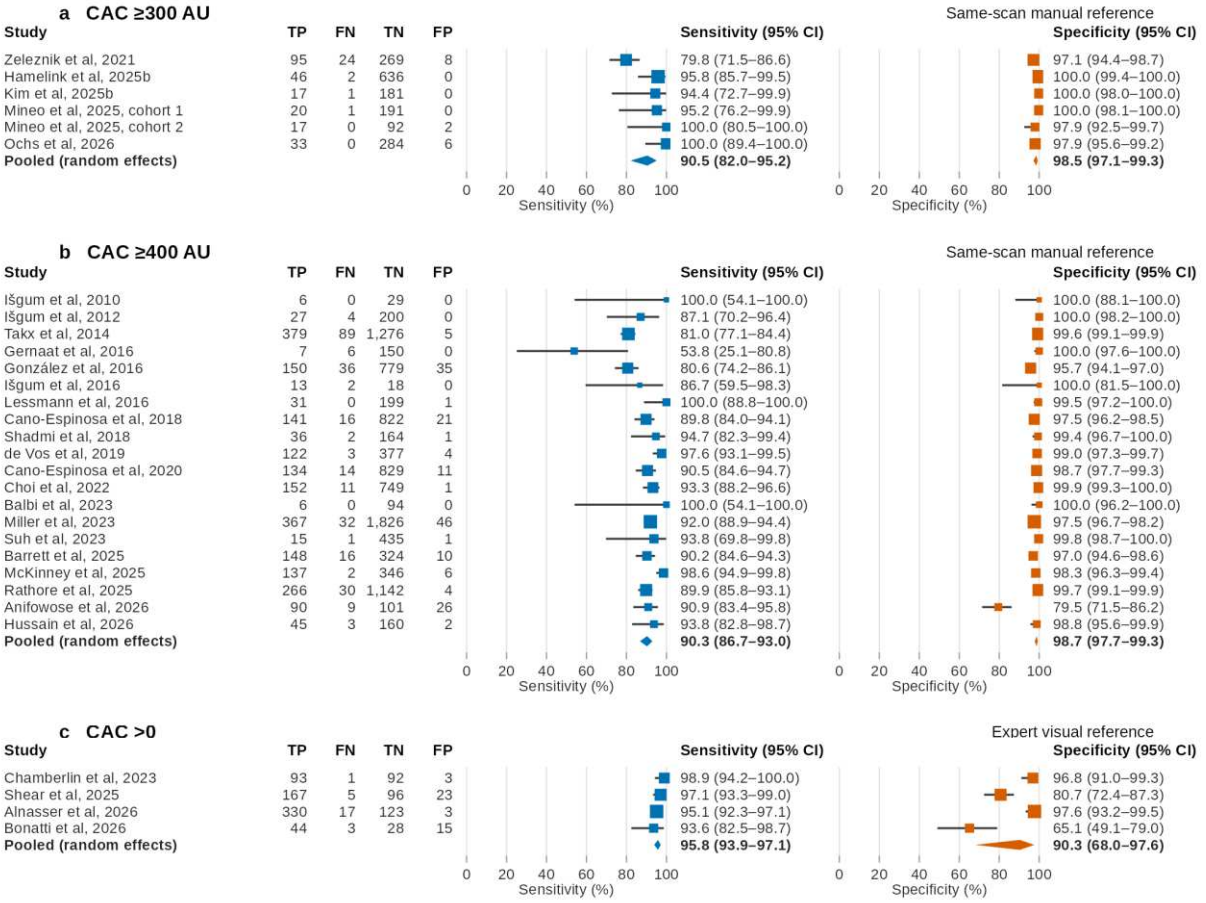


Figure 6. Sensitivity and specificity of AI-derived CAC against manual scoring of the same non-gated CT at CAC ≥300 and ≥400 AU, and against expert visual assessment. a, CAC ≥300 AU (six estimates from five reports; 1,925 participants). **b**, CAC ≥400 AU (20 estimates from 20 reports; 12,742 participants). **c**, Expert visual assessment of CAC presence (four estimates from four reports; 1,043 participants). Plotting conventions as in Fig. 3.

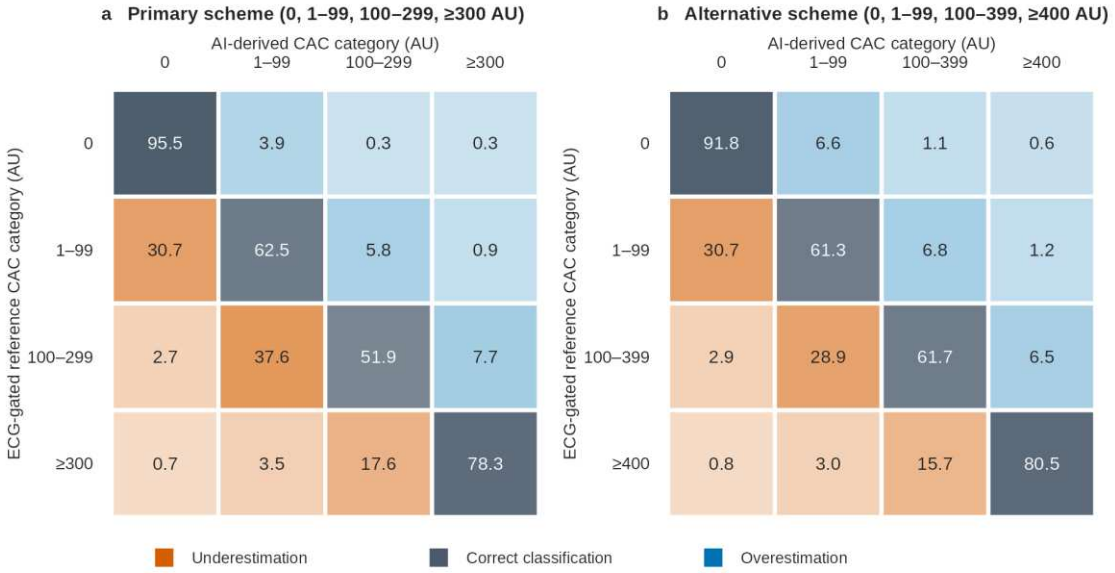


Figure 7. Cross-classification of AI-derived and ECG-gated CAC categories. Posterior mean probabilities (%) of each AI-derived category within each ECG-gated reference category for **a**, the primary scheme of 0, 1–99, 100–299, and ≥300 AU (six tables from four reports; 1,727 participants) and **b**, the alternative scheme of 0, 1–99, 100–399, and ≥400 AU (20 tables from 15 reports; 7,935 participants). Rows are ECG-gated reference categories and columns AI-derived categories on paired non-gated CT; each row sums to 100% before rounding. The reference categories comprised 649, 422, 328, and 328 participants in the primary scheme and 2,418, 1,943, 1,576, and 1,998 participants in the alternative scheme. Orange cells indicate underestimation, grey cells correct classification, and blue cells overestimation; darker shading denotes higher probability within each colour. The two schemes were fitted separately with no shared reports. Posterior estimates and 95% credible intervals are given in Supplementary Tables 8 and 9.

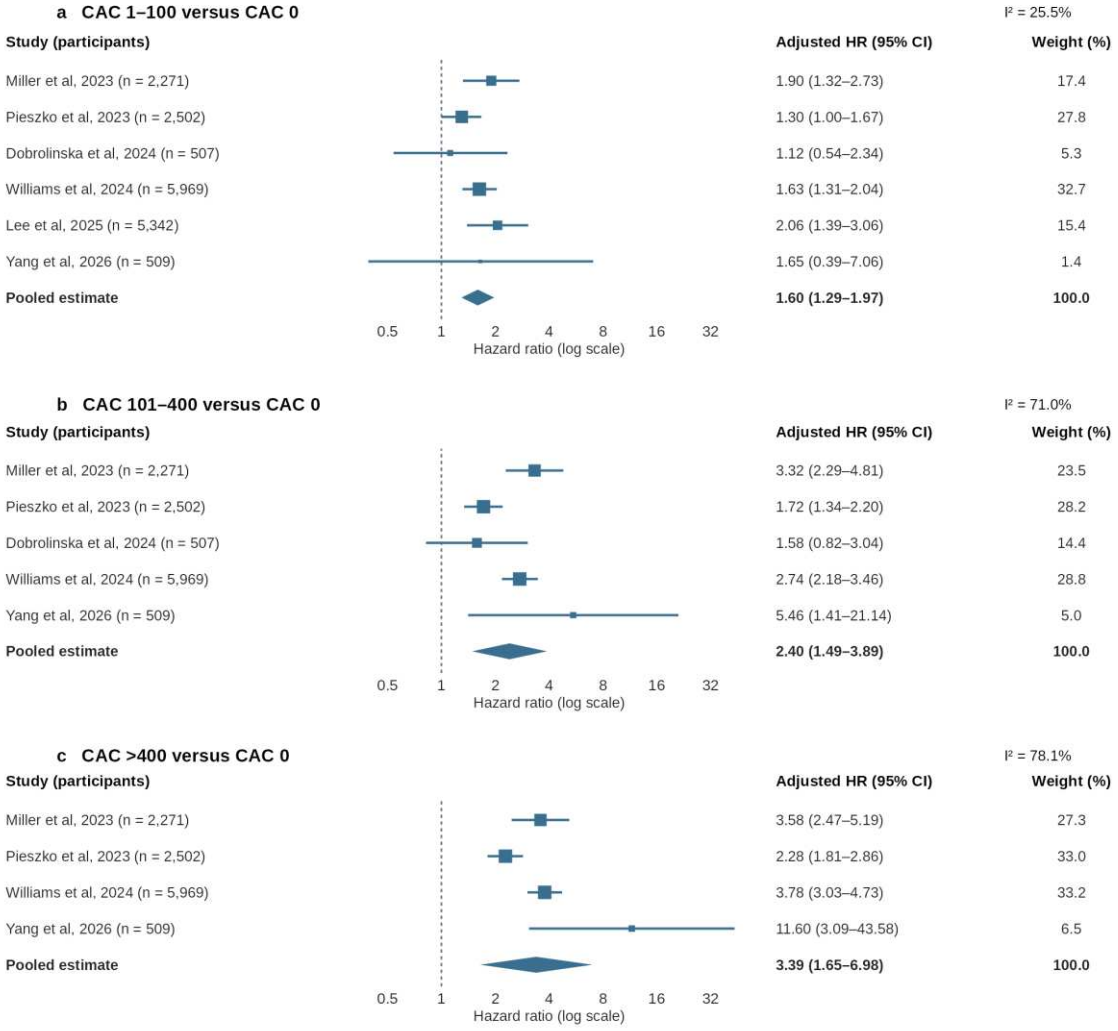


Figure 8. Association of AI-derived CAC with major adverse cardiovascular events. Study-specific adjusted hazard ratios compare AI-derived CAC **a**, 1–100 (six reports; 17,100 participants); **b**, 101–400 (five reports; 11,758 participants); and **c**, >400 (four reports; 11,251 participants) with CAC 0. Squares show study estimates, with area proportional to random-effects weight; horizontal lines show 95% confidence intervals. Diamonds show pooled estimates from random-effects models with 95% confidence intervals; I^2 is the between-study heterogeneity. Participant totals are sums of report-specific analysis populations and may include the same individuals in more than one contrast. MACE definitions and adjustment covariates are detailed in Supplementary Table 14. HR, hazard ratio.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 04SupplementaryData1.xlsx
- 03SupplementaryInformation.pdf