AI in medicine

MDPI

Review

Artificial Intelligence for Coronary Artery Disease Prediction Using ECG and CCTA: A Systematic Review

Ahmad Ibrahim Alshdaifat [1,2], Wamadeva Balachandran [1,*] and Ziad Hunaiti [1]

[1] Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK; ahmad.alshdaifat@brunel.ac.uk or ah.sh_97@yahoo.com (A.I.A.); ziad.hunaiti@gmail.com (Z.H.)
[2] College of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan
* Correspondence: wamadeva.balachandran@brunel.ac.uk

Abstract

Coronary artery disease (CAD) is the leading cause of death worldwide, highlighting the need for more reliable and efficient diagnostic tools beyond conventional methods. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has shown strong potential for detecting obstructive CAD by learning complex patterns from electrocardiogram (ECG) and coronary computed tomography angiography (CCTA) data. This rapid systematic review assesses and compares the diagnostic performance and methodological quality of AI models built for CAD prediction using ECG and CCTA data. A systematic search following PRISMA-ScR guidelines was conducted for primary studies published between 2021 and 2025. Eleven studies were included, six using ECG data and five using CCTA data. Methodological quality was evaluated using the PROBAST+AI tool. ECG-based models achieved AUCs of 0.72–0.961 and CCTA-based models showed slightly stronger top-end performance, with AUCs of 0.77–0.97. External validation was uncommon in both groups, applied in only 40% of CCTA studies and 33% of ECG studies, so neither modality demonstrated clearly greater validation maturity. Despite these strong results, PROBAST+AI assessment revealed a high risk of bias in 90.9% of the included studies, largely due to weaknesses in the analysis domain, including poor handling of missing data and the absence of model calibration reporting. AI models show strong diagnostic accuracy for CAD across both modalities, although external validation was limited and applied in a minority of studies. However, the widespread methodological bias means these tools should currently support clinical decision-making rather than replace standard diagnostic methods. Future studies should focus on prospective multicentre validation and the use of multimodal data.

Keywords: artificial intelligence; coronary artery disease; deep learning; electrocardiogram; coronary computed tomography angiography; machine learning; systematic review; PROBAST

Academic Editor: Daniel B. Hier

Received: 23 April 2026
Revised: 28 July 2026
Accepted: 5 August 2026
Published: 13 August 2026

1. Introduction

Coronary artery disease (CAD) is the leading cause of death and illness worldwide, responsible for an estimated 17.8 million deaths each year [1]. It accounts for over 31% of all global deaths, making it a major public health concern [2]. CAD represents a major economic burden in the UK, with cardiovascular disease overall costing the economy £29 billion in 2021/22, of which CAD accounted for a substantial share of direct healthcare expenditure [3]. Although invasive coronary angiography (CAG) remains the gold standard

for definitive diagnosis, its routine use is limited by high costs and procedural risks [4]. The limitations of current diagnostic pathways are further shown by low diagnostic yields, as only 37.6% of patients referred for elective angiography are found to have obstructive disease [5]. In addition, common non-invasive screening tools, such as the standard 12-lead electrocardiogram (ECG), show limited accuracy for detecting stable obstructive disease, with Sensitivity of 0.515 and Specificity of 0.661 [6]. While more advanced imaging tools like coronary computed tomography angiography (CCTA) offer better anatomical detail, they are often affected by subjective interpretation and a known tendency to overestimate luminal stenosis [7]. Therefore, there is a clear clinical need for more accurate, efficient, and accessible diagnostic approaches to improve patient risk assessment and resource use [8].

Artificial intelligence (AI), including both machine learning (ML) and deep learning (DL), is changing the landscape of clinical diagnostics by enabling the detection of complex, nonlinear patterns in large datasets [2]. DL is an advanced branch of AI that uses multilayered neural networks to learn directly from raw data without requiring manually engineered features [9]. These computational models are well suited to cardiovascular diagnostics, where ECG signals and high resolution CCTA images provide rich data for AI analysis [10]. For the 12-lead ECG, AI algorithms can detect subtle electrical changes that are not visible to the human eye, greatly improving the diagnostic value of this widely available screening tool [11]. This effectively transforms the resting ECG from a low accuracy test into a powerful digital biomarker capable of identifying significant coronary lesions and disease burden [12]. At the same time, AI methods address the complexity of CCTA by automating time consuming tasks such as coronary tree reconstruction, plaque detection, and vessel segmentation [7]. As argued in [13], this automation improves diagnostic consistency by reducing subjective variability and lowering the high computational demands typically associated with advanced image analysis. Overall, integrating AI into these modalities offers a practical way to close the existing diagnostic gap by providing more objective, precise, and efficient risk stratification [4].

A growing body of primary research has documented the development of AI models for CAD detection using ECG data [4], while parallel studies have also shown the strong diagnostic potential of AI-enhanced CCTA in identifying functionally significant coronary lesions [7]. Despite this rapid growth in primary evidence, the field remains fragmented and lacks the comprehensive synthesis needed to guide evidence-based clinical practice [11]. A key gap in the existing literature is the absence of a direct comparison between AI models developed from fundamentally different modalities, such as the 12-lead ECG and detailed CCTA imaging [14]. Furthermore, the methodological quality and risk of bias in these individual studies have not been thoroughly assessed using dedicated appraisal tools [12]. Importantly, the clinical and economic trade-offs between a low cost, scalable AI ECG screening approach and a resource-intensive AI CCTA anatomical assessment remain poorly understood for real world clinical use [6]. This lack of structured evidence synthesis is a major barrier to progress, limiting the effective integration of AI diagnostics into routine cardiovascular care [13].

Therefore, this rapid systematic review aims to comprehensively synthesise and critically compare machine learning and deep learning models developed for CAD prediction using ECG and CCTA data. To achieve this, the review addresses three main objectives. First, it synthesises the reported diagnostic performance of models from each modality, including sensitivity, specificity, and discriminatory capacity. Second, it provides a critical comparison of ECG and CCTA-based models, examining their relative predictive value and clinical applicability across different cardiovascular care settings. Third, it formally evaluates the methodological quality and risk of bias in the included studies using the

PROBAST+AI assessment tool. Using a structured PRISMA-guided approach and narrative synthesis, this review offers a clear appraisal of the current state of AI in cardiovascular diagnostics to support its evidence-based clinical translation. To address these objectives, this review is guided by the following research questions:

- RQ1: What diagnostic performance do AI models achieve for CAD prediction using ECG and CCTA data individually?
- RQ2: How do ECG-based and CCTA-based AI models compare in terms of predictive performance, validation approach, and clinical applicability?
- RQ3: What are the key sources of bias and quality concerns limiting the clinical translation of AI-based CAD prediction models?

2. Methods

2.1. Study Design

This rapid systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines. The review was designed as a rapid systematic review, meaning several components of the standard systematic review process were deliberately streamlined to produce a timely synthesis of a fast-moving field, while preserving methodological transparency. In practice, this meant applying a focused publication window (2021–2025), a title-based search strategy, a restricted set of three major databases, an English-language restriction, and the exclusion of conference abstracts and other non-primary publication types, with screening and data extraction carried out by a single reviewer. The implications of these design choices for the completeness of study retrieval are discussed further in the limitations. Before starting the literature search, a detailed internal protocol was developed to predefine the research questions, eligibility criteria using the PICOS framework, search strategy, data extraction process, and analysis plan. This was done to reduce bias and ensure methodological transparency throughout the review. The key components of this protocol are described in the following sections of this manuscript. This review was prospectively registered on the Open Science Framework https://doi.org/10.17605/OSF.IO/H793Y.

2.2. Eligibility Criteria

Study eligibility was defined using the PICOS framework:

- P (Population): Patients with suspected or known coronary artery disease.
- I (Intervention): Artificial intelligence algorithms, including machine learning and deep learning, for CAD prediction.
- C (Comparison): Standard clinical assessment and diagnostic methods.
- O (Outcome): Diagnostic and prognostic accuracy metrics including AUC, Sensitivity, and Specificity.
- S (Study Design): Primary diagnostic accuracy or prognostic studies.

Based on this framework, the following inclusion and exclusion criteria were applied.

2.2.1. Inclusion Criteria

This review included primary research articles that developed or validated a machine learning or deep learning model for CAD prediction, where the input data came solely from either an electrocardiogram (ECG) or coronary computed tomography angiography (CCTA). The study population was limited to adult patients (≥18 years) with suspected or known CAD. Articles were required to report at least one quantitative diagnostic or prognostic accuracy metric, specifically the area under the receiver operating characteristic curve (AUC), to allow for performance evaluation. Only full-text articles published in

English between 2021 and 2025 were considered, ensuring the review focused on the most recent and relevant advances in the field.

2.2.2. Exclusion Criteria

A wide range of publication types was excluded to maintain focus on primary applicable research. These included duplicate records, conference papers, letters, books, book chapters, published errata, case reports, clinical trial protocols, comments, and editorials. Articles published outside the specified date range and those not in English were also excluded. In addition, studies were excluded if they involved patient cohorts with specific non-CAD primary conditions, such as isolated cardiomyopathy or post cardiac transplant populations, to ensure the findings were applicable to the target population of suspected CAD. Studies using AI models based on data modalities other than ECG or CCTA, such as echocardiography alone or purely clinical scores, were excluded, as were articles that did not report the required performance metrics including AUC, Sensitivity, and Specificity.

2.3. Information Sources and Search Strategy

A systematic and comprehensive literature search was conducted in December 2025 to identify all relevant primary research. The search was carried out across three major electronic databases known for their broad coverage of biomedical, interdisciplinary, and engineering literature: PubMed, Web of Science Core Collection, and Scopus.

The search strategy was designed to capture studies at the intersection of coronary artery disease, artificial intelligence, and diagnostic prediction. Controlled vocabulary and keywords were combined to target the population ("Coronary Artery Disease"), the intervention ("Machine Learning," "Deep Learning"), and the application ("predict," "diagnose"). The search was kept intentionally broad to maximise sensitivity, with date restrictions (2021–2025) and publication type exclusions applied during the screening phase based on the eligibility criteria. The search syntax was adapted to each database using the appropriate field tags and Boolean operators, as shown below:

- PubMed: ((("Coronary Artery Disease"[Title] OR "CAD"[Title])) AND (("Artificial Intelligence"[Title] OR "Machine Learning"[Title] OR "Deep Learning"[Title])) AND ((predict*[Title] OR diagnos*[Title] OR detection[Title]))) NOT (review[Publication Type])
- Web of Science: (TI = ("Coronary Artery Disease" OR "CAD") AND TI = ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND TI = (predict* OR diagnos* OR detection)) NOT (DT = (Review))
- Scopus: (TITLE ("Coronary Artery Disease" OR "CAD") AND TITLE ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND TITLE (predict* OR diagnos* OR detection)) AND NOT (DOCTYPE ("re"))

2.4. Study Selection Process

The study selection process was carried out in line with the PRISMA-ScR guidelines (Table S1). All records retrieved from the database searches were collected and managed using the systematic review software Rayyan [15] (https://www.rayyan.ai, accessed on 15 December 2025). This platform was used to automatically identify and remove duplicate entries, followed by a manual check to confirm that no duplicates remained. The subsequent screening was performed by a single reviewer. The process began with an initial screening of titles and abstracts, where each record was assessed against the eligibility criteria. Records that appeared relevant or where relevance was uncertain moved forward to a full-text review for final eligibility determination. To ensure methodological rigour and reduce the risk of bias associated with a single-reviewer process, a systematic protocol was followed throughout. Any uncertainties regarding eligibility were flagged and re-evaluated in a

dedicated session, with final decisions made strictly against the predefined PICOS criteria. A detailed log was kept recording the reasons for exclusion for every article assessed at the full-text stage. The complete results of this selection process are shown in the PRISMA flow diagram (Figure 1), which documents the number of studies identified, screened, evaluated, and included, along with the reasons for exclusion at the full-text assessment stage.

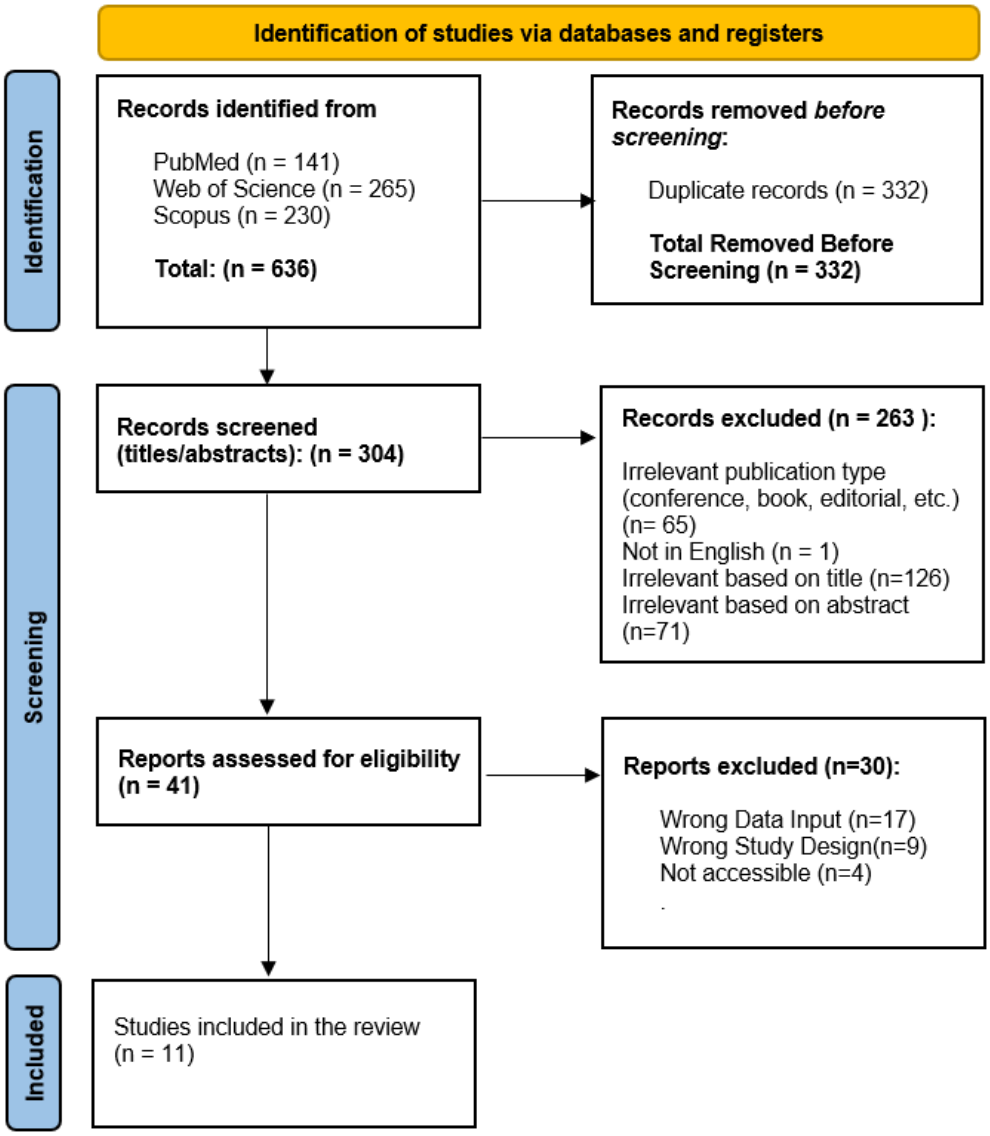

Figure 1. PRISMA flow diagram of study selection process.

Although dual independent review is recommended to reduce bias, screening, data extraction, and quality assessment were conducted by a single reviewer. To mitigate this, all stages followed the predefined protocol and PICOS criteria, uncertain cases were revisited in a dedicated session, and a full-text exclusion log was kept throughout. The potential effect of this limitation on study retrieval, inclusion, extraction, and risk-of-bias assessment is acknowledged and factored into the interpretation of findings.

2.5. Data Extraction Process

A structured data extraction framework was applied following the final inclusion of studies. Given the focus on qualitative synthesis and comparative analysis, a standardised extraction protocol was carried out manually using a set of customised tables developed in

Microsoft Word. This approach was designed to systematically capture and organise the varied data relevant to the review objectives.

The extraction protocol was organised into four complementary tables, each capturing a distinct area of information from the included articles. Table 1 (Identification and Study Scope) provides a foundational overview, recording bibliographic details, study design, geographical origin, clinical focus, and primary data modality (ECG or CCTA). Table 2 (Patient and Data Characteristics) documents the scale and nature of the input data, including total sample size, number of records, the specific format of ECG signals or CCTA images, and any additional clinical variables used. Table 3 (Model Architecture and Technical Framework) details the computational methodology, recording the type of AI approach used (ML, DL, or hybrid), the specific algorithms or architectures employed, and the validation strategy. Finally, Table 4 (Performance Metrics) collates the quantitative diagnostic accuracy outcomes, including AUC, ACC, Sensitivity, Specificity, and F1 Score, along with their corresponding confidence intervals where reported.

Table 1. Identification and study scope of included studies.

Authors, Year	Study Design/Type	Country	Clinical Focus	Data
Yin-Hao Lee et al., 2023 [4]	Retrospective Multicentre	Taiwan	Obstructive Coronary Artery	ECG
Leasure et al., 2021 [6]	Retrospective Single-centre	USA	A: Multiclass CAD B: Binary Classification	ECG
Hyun-Gyu Lee et al., 2023 [16]	Retrospective Single-centre	South Korea	Obstructive Coronary Artery	ECG
Huang et al., 2022 [17]	Retrospective Multicentre	Taiwan	Significant Coronary Artery	ECG
Park et al., 2023 [18]	Retrospective Single-centre	Republic of Korea	Obstructive Coronary Artery	ECG
Yeh et al., 2025 [11]	Retrospective Multicentre	Taiwan	Significant Coronary Artery	ECG
Guo et al., 2024 [7]	Retrospective Multicentre	China	Significant Coronary Artery	CCTA
Kigka et al., 2022 [14]	Prospective Multicentre	Europe	Obstructive Coronary Artery	CCTA
Wahab Sait and Dutta, 2023 [13]	Retrospective Multicentre	China	CAD	CCTA
Yang et al., 2023 [19]	Retrospective Multicentre	China	Obstructive Coronary Artery	CCTA
Kim et al., 2025 [20]	Retrospective Multicentre	Republic of Korea	Obstructive Coronary Artery	CCTA

Table 2. Patient and data characteristics of included studies.

Authors, Year	Total Patients	Total Records	ECG/CCTA Data	Other Data
Yin-Hao Lee et al., 2023 [4]	4951	4951	12-lead ECGs	CVRFs
Leasure et al., 2021 [6]	1659	1659	12-lead ECGs	N/A
Hyun-Gyu Lee et al., 2023 [16]	7907	7907	12-lead ECGs	Clinical and laboratory
Huang et al., 2022 [17]	3356	15,044	12-lead ECGs	N/A
Park et al., 2023 [18]	723	723	Image-based ECG data	N/A
Yeh et al., 2025 [11]	5202	5376	Raw ECG XML data	N/A
Guo et al., 2024 [7]	463	2144	CCTA images	N/A
Kigka et al., 2022 [14]	187	287	CCTA images	Demographics, risk factors, and 20+ biohumoral markers (LDL, NT-proBNP, TSH, etc.)
Wahab Sait and Dutta, 2023 [13]	Dataset 1: 500 Dataset 2: 1000	Dataset 1: 2364 images Dataset 2: 1000 3D CCTA	CCTA images	N/A
Yang et al., 2023 [19]	374	1495	CCTA images	N/A
Kim et al., 2025 [20]	676	10,060	Curved multiplanar Reconstruction (MPR) images	N/A

Note: N/A: not applicable; CVRFs: cardiovascular risk factors.

Table 3. Model architecture and technical framework of included studies.

Authors, Year	Modality	ML/DL	Algorithms	Validation	Notes
Yin-Hao Lee et al., 2023 [4]	ECG	Both	Elastic Net (EN) + 1D CNN	Both	Stacking method
Leasure et al., 2021 [6]	ECG	DL	ECGio	Internal	–
Hyun-Gyu Lee et al., 2023 [16]	ECG	Both	(LR, LightGBM, XGBoost) + ResNet-1D-CNN	Internal	Ensemble
Huang et al., 2022 [17]	ECG	DL	InceptionV3 CNN	Internal	–
Park et al., 2023 [18]	ECG	Both	ResNet-based CNN, XGBoost	Internal	Tree-based ensemble
Yeh et al., 2025 [11]	ECG	ML	XGBoost	External	–
Guo et al., 2024 [7]	CCTA	Both	Modified U-Net, Reduced-Order 3D CFD	Internal Cohorts	–
Kigka et al., 2022 [14]	CCTA	ML	Gradient Boosting Classifier, SVM with Recursive Feature Elimination (RFE)	Internal	–
Wahab Sait and Dutta, 2023 [13]	CCTA	DL	UNet++ model, YOLOv7	Internal	–
Yang et al., 2023 [19]	CCTA	DL	Dual-head multi-task ResNet	External	–
Kim et al., 2025 [20]	CCTA	DL	YOLOv4	External	–

Abbreviations: ECG—electrocardiogram; CCTA—coronary computed tomography angiography; ML—machine learning; DL—deep learning; CNN—convolutional neural network; LR—logistic regression; XGBoost—eXtreme Gradient Boosting; SVM—support vector machine; RFE—recursive feature elimination; CFD—computational fluid dynamics. "Both" indicates the study combined traditional ML and DL components (e.g., in a hybrid or ensemble pipeline). Validation: Internal = model trained and evaluated using data drawn from the same cohort/source (e.g., holdout split or cross-validation); External = model evaluated on an independent cohort or dataset separate from training data. Note: The model characteristics and validation approaches reported here come from independent studies that differ considerably in their design and methods, and should not be read as direct head-to-head comparisons between ECG-based and CCTA-based models. Differences seen across studies may well reflect variation in patient populations, CAD outcome definitions, reference standards, disease prevalence, and data quality, rather than any genuine difference in how the two modalities perform.

Table 4. Performance metrics of included studies.

Authors, Year	Modality	Validation Source	AUC	ACC	Sens.	Spec.	F1
Yin-Hao Lee et al., 2023 [4]	ECG	Internal	0.72 (95% CI: 0.68–0.77)	0.69	0.75	0.62	0.72
Leasure et al., 2021 [6] (A)	ECG	Internal	N/A	0.957	0.951	0.967	0.933
Leasure et al., 2021 [6] (B)	ECG	Internal	0.961 (95% CI: 0.936–1.00)	0.899	0.930	0.856	0.915
Hyun-Gyu Lee et al., 2023 [16]	ECG	Internal	0.767	N/A	0.761	0.625	0.696
Huang et al., 2022 [17] (InceptionV3)	ECG	Internal	0.869	0.900 ± 0.012	0.899	0.903	N/A
Park et al., 2023 [18]	ECG	Internal	0.802	0.757	0.758	0.756	N/A
Yeh et al., 2025 [11]	ECG	External	0.91	0.77	0.84	N/A	N/A
Guo et al., 2024 [7]	CCTA	Internal	0.82	0.82	0.84	0.81	N/A
Kigka et al., 2022 [14]	CCTA	Internal	0.82 ± 0.11	0.81 ± 0.08	0.88 ± 0.15	0.73 ± 0.09	N/A
Wahab Sait and Dutta, 2023 [13] (D1)	CCTA	Internal	0.97	0.994	0.987	0.987	0.986
Wahab Sait and Dutta, 2023 [13] (D2)	CCTA	Internal	0.96	0.995	0.990	0.993	0.990
Yang et al., 2023 [19]	CCTA	External	0.77	N/A	0.936	0.596	N/A
Kim et al., 2025 [20]	CCTA	External	0.871	N/A	0.933	0.807	N/A

Abbreviations: ECG—electrocardiogram; CCTA—coronary computed tomography angiography; AUC—area under the receiver operating characteristic curve; ACC—accuracy; CI—confidence interval; F1—F1 score (harmonic mean of precision and recall); N/A—not reported by the original study. Sens.—Sensitivity (Recall); Spec.—Specificity. Validation Source: Internal—metric derived from a holdout split or cross-validation within the same cohort/dataset used for model development; External—metric derived from an independent cohort, dataset, or site not used in model training. (A) and (B) refer to the two analyses reported by Leasure et al. [6]: (A) is the study's primary analysis, a three-class (multiclass) classification of stenosis severity (no-to-mild, moderate, severe:

based on maximum percent diameter stenosis); (B) is the secondary analysis, a binary classification of significant versus non-significant disease at a ≥50% stenosis threshold. (D1) and (D2) refer to the two independent datasets used by Wahab Sait and Dutta [13] to evaluate the same model: (D1)—Dataset 1 (500 patients; 2D straightened-vessel CCTA images from a public repository); (D2)—Dataset 2 (1000 patients; 3D CCTA volumes from a single hospital cohort). The performance metrics summarised above derive from separate, non-comparable study cohorts, outcome definitions, and validation designs. Differences in AUC, accuracy, sensitivity, or specificity across studies do not in themselves indicate that one modality outperforms the other (see also Sections 3.4.3 and 4.2).

2.6. Quality and Risk of Bias Assessment

The methodological quality and risk of bias of the included studies were assessed using the PROBAST+AI tool (Prediction model Risk Of Bias ASsessment Tool updated for Artificial Intelligence methods), which is specifically designed to evaluate prediction models developed using either traditional regression or artificial intelligence methods. This tool includes additional considerations relevant to AI-based models and supports a structured evaluation across multiple domains.

The assessment began with defining the intended purpose of the prediction models under review using the PICOS framework, as described in the Eligibility Criteria subsection. Each included study was then classified according to its developmental stage. All studies in this review involved both model development and evaluation, meeting the criteria for combination studies as defined by PROBAST+AI.

Following this classification, a domain-based assessment was carried out for both the model development and model evaluation components of each study. Four domains were examined: participants and data sources, which addressed participant selection and the appropriateness of data sources; predictors, which considered the definition, measurement, and timing of ECG or CCTA input data; outcome, which evaluated the definition, measurement, and timing of CAD diagnosis; and analysis, which examined key methodological aspects including sample size adequacy, handling of missing data, mitigation of overfitting, and the appropriateness of performance measures.

For the model development component, each domain was rated for quality concerns as low, high, or unclear. For the model evaluation component, which covered validation of model performance, each domain was assessed for risk of bias using the same rating scale. The first three domains were also evaluated for applicability concerns relative to this review's predefined research question and intended context of use.

An overall judgement was then made for both development quality and evaluation risk of bias according to PROBAST+AI criteria. Studies were assigned low concern or low risk if all domains received low ratings, high concern or high risk if any single domain was rated high, and unclear concern or risk if at least one domain was rated unclear with no high ratings present. The assessment was conducted by a single reviewer, with consistent application of the tool ensured by referring to the official PROBAST+AI Explanation and Elaboration Light guide throughout (https://www.probast.org, accessed on 15 January 2026). The complete results of this quality assessment are presented and summarised in tabular and graphical format (Figures 2–5).

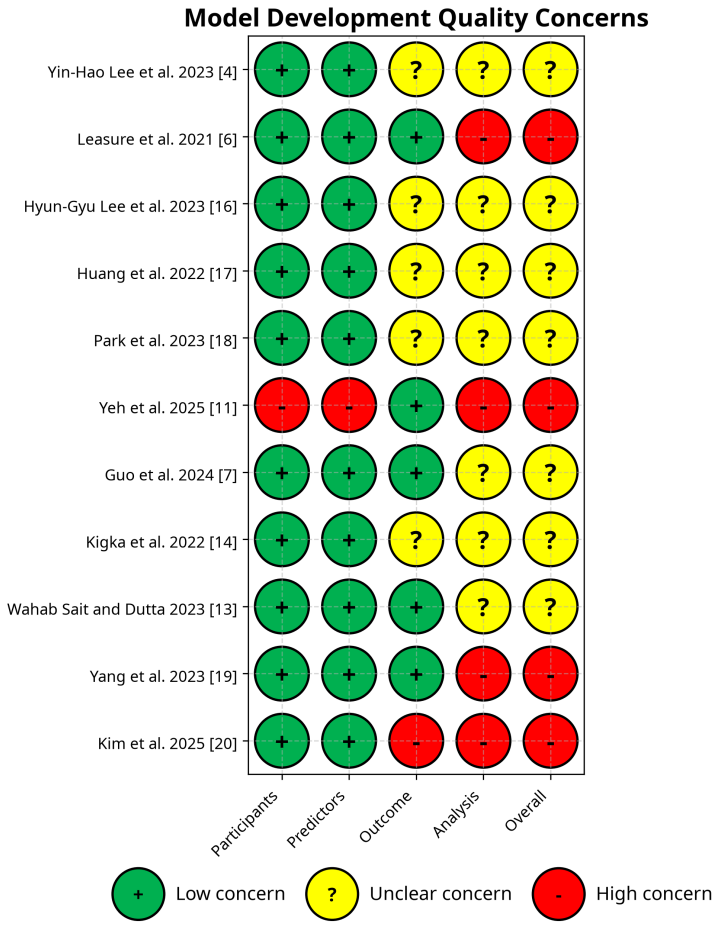

Figure 2. PROBAST+AI assessment of model development quality concerns across all included studies.

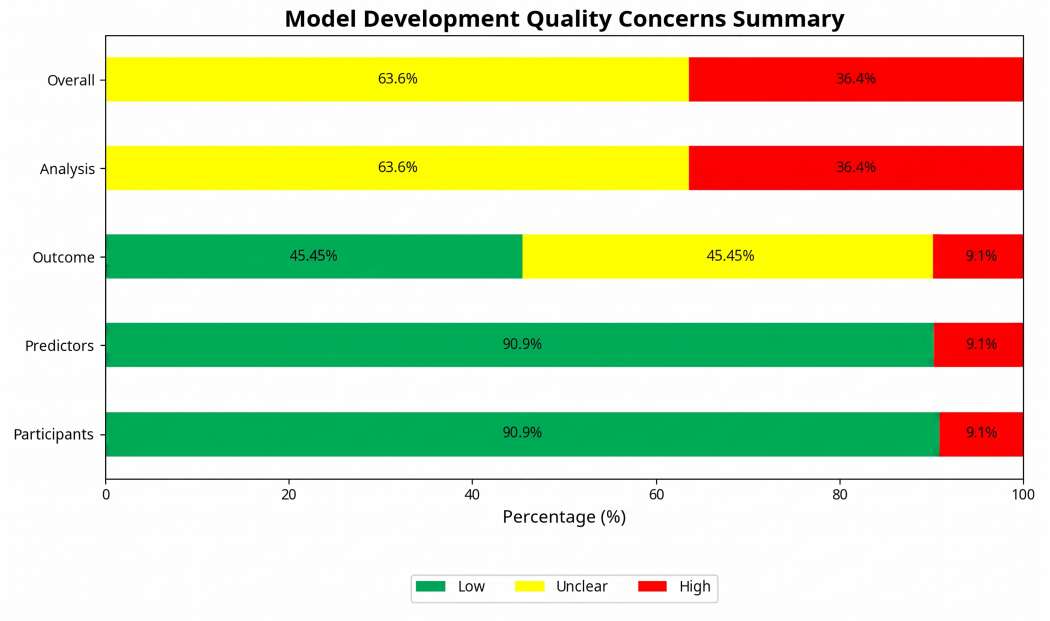

Figure 3. PROBAST+AI assessment of model development quality concerns summary.

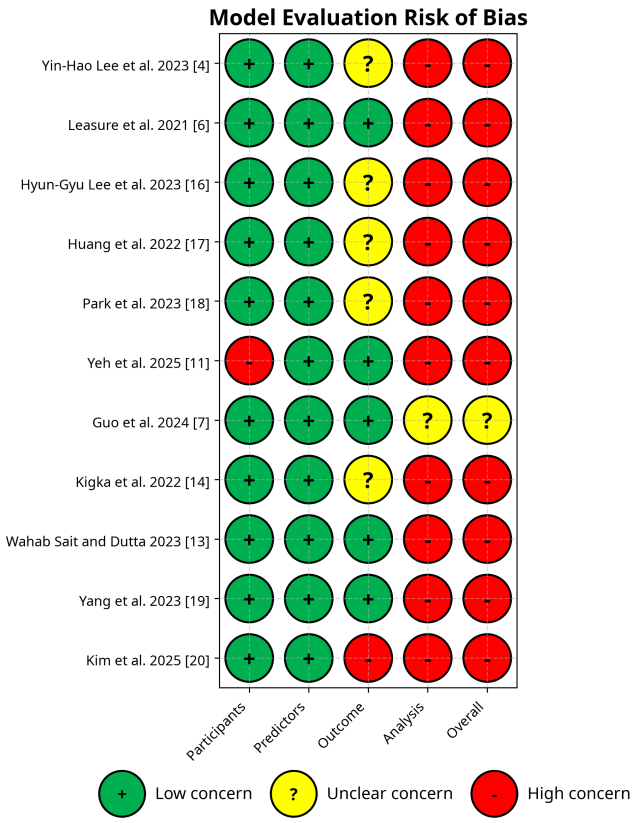

Figure 4. PROBAST+AI assessment of model evaluation risk of bias across all included studies.

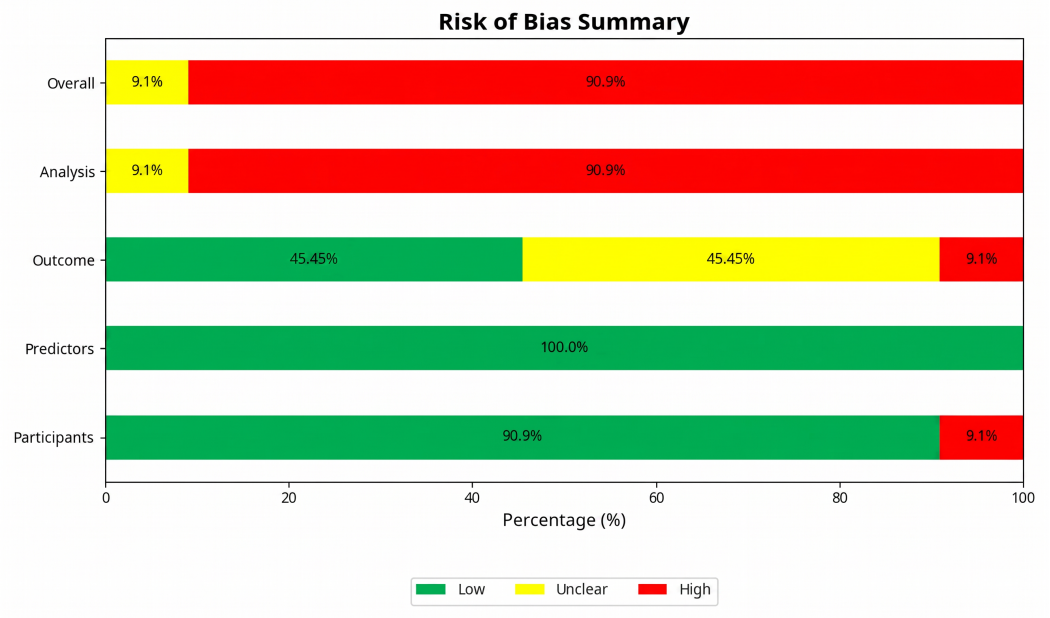

Figure 5. PROBAST+AI assessment of model evaluation risk of bias summary.

3. Results

3.1. Study Selection

The systematic search of PubMed, Web of Science, and Scopus databases retrieved a total of 636 records. After removing 332 duplicate records, 304 unique records remained for title and abstract screening.

During the title and abstract screening phase, 263 records were excluded for the following reasons: 65 were irrelevant publication types such as conference papers, books,

and editorials; one was not published in English; 126 were irrelevant based on title; and 71 were irrelevant based on abstract. This left 41 full-text articles to be retrieved and assessed for eligibility.

Of these 41 full-text articles, 30 were excluded. The main reason for exclusion was the use of AI models based on data modalities other than ECG or CCTA, such as echocardiography or clinical scores alone, which accounted for 17 exclusions. A further nine articles were excluded due to inappropriate study design, including editorials or studies without original model development or validation. Four additional articles were inaccessible in full text despite retrieval attempts and were therefore excluded.

A total of 11 studies met all inclusion criteria and were included in this systematic review. Of these, six studies developed and evaluated AI models using ECG data, and five used CCTA data. The complete study selection process is illustrated in the PRISMA-ScR flow diagram presented in Figure 1.

3.2. Study Characteristics

The characteristics of the 11 included studies are summarised in Table 1 (Identification and Study Scope) and Table 2 (Patient and Data Characteristics). The studies were published between 2021 and 2025 and came from diverse geographical regions, including Taiwan, the United States, South Korea, China, and a multicentre European collaboration.

Regarding study design, the majority used a retrospective design, with seven being retrospective multicentre and three being retrospective single-centre. One study was a prospective multicentre study, and no prospective single-centre studies were identified. Sample sizes ranged widely from 187 to 7907 patients, with the total number of records ranging from 287 to 15,044, reflecting variation in cohort availability and data collection scale. In several studies the number of records exceeded the number of patients because multiple recordings were included per individual. In Huang et al. [17], for example, the 15,044 records correspond to all available 12-lead ECGs collected per patient within the defined clinical window, comprising every ECG recorded in the 30 days before percutaneous coronary intervention for patients with CAD and multiple outpatient recordings for those without, rather than a single ECG per patient.

Among the six ECG-based studies, the standard 12-lead electrocardiogram was the primary data format used in four studies. The remaining two ECG studies used image-based ECG data and raw ECG XML data. Among the five CCTA-based studies, four used standard CCTA images, while one study used Curved Multiplanar Reconstruction (MPR) images as the input data modality.

All studies focused on the prediction of obstructive or significant coronary artery disease. The clinical outcome was typically defined using reference standards such as invasive coronary angiography with stenosis thresholds of $\geq$50% or $\geq$70%, or expert interpretation of CCTA. While most studies relied solely on ECG or CCTA data, three studies incorporated additional clinical variables. One ECG study included cardiovascular risk factors, another combined clinical and laboratory data, and one CCTA study used a comprehensive set of demographics, risk factors, and over 20 biohumoral markers, including LDL, NT-proBNP, and TSH. The remaining eight studies developed models based solely on the primary imaging or signal data.

3.3. Risk of Bias and Quality Assessment

The methodological quality and risk of bias of the 11 included studies were assessed using the PROBAST+AI tool, with separate evaluations for model development and model evaluation components. The complete results are presented in Figures 2–5.

3.3.1. Model Development Quality Concerns

For the model development component, the overall quality concern was high in four studies (36.4%), unclear in seven studies (63.6%), and low in none (Figures 2 and 3). Across the individual domains, the Participants domain showed low concern in ten studies (90.9%) and high concern in one study (9.1%). The Predictors domain similarly showed low concern in ten studies (90.9%) and high concern in one study (9.1%). The Outcome domain showed greater variability, with five studies rated as low concern (45.45%), five as unclear (45.45%), and one as high concern (9.1%). The Analysis domain showed no studies achieving low concern, seven studies rated as unclear (63.6%), and four rated as high concern (36.4%). These results reflect common methodological issues, including inadequate sample size relative to model complexity and a lack of clarity regarding methods used to control for overfitting during model training.

3.3.2. Risk of Bias–Model Evaluation

For the model evaluation component, the overall risk of bias was high in ten studies (90.9%), unclear in one study (9.1%), and low in none (Figures 4 and 5). Across the individual domains, the Participants domain showed low risk in ten studies (90.9%) and high risk in one study (9.1%). The Predictors domain showed the strongest results, with all 11 studies rated as low risk (100.0%). The Outcome domain showed mixed results, with five studies rated as low risk (45.45%), five as unclear (45.45%), and one as high risk (9.1%). The Analysis domain was the main driver of the high overall risk-of-bias ratings, with no studies achieving low risk, one study rated as unclear (9.1%), and ten rated as high risk (90.9%). This reflects widespread methodological weaknesses in validation practices, including reliance on internal validation only, potential data leakage between training and test sets, and the absence of reported confidence intervals for performance metrics.

3.3.3. Applicability Concerns

Regarding applicability to the review's intended purpose as defined in the PICOS framework, concerns were low across all domains. The Participants, Predictors, and Outcome domains all showed low applicability concerns, indicating that the study populations, predictors, and outcome definitions were well aligned with the review's focus on adult patients with suspected CAD undergoing ECG or CCTA evaluation.

While the included studies demonstrated strong quality in participant selection and predictor handling, notable concerns emerged in the analysis and validation phases. The high risk of bias in the evaluation component across 90.9% of studies, combined with the absence of any study achieving low overall risk, suggests that the reported performance metrics for these AI models should be interpreted with caution. These findings highlight a clear need for more rigorous validation practices, particularly external validation on independent datasets, in future studies developing AI models for CAD prediction using ECG and CCTA data.

3.4. Synthesis of Results

The model architectures, validation approaches, and performance metrics for all 11 studies are detailed in Table 3 (Model Architecture and Technical Framework) and Table 4 (Performance Metrics). A concise summary of the principal methodological strengths and weaknesses of each included study is presented in Table 5.

Table 5. Summary of key methodological strengths and weaknesses of the included studies.

Study (Year)	Modality	Validation	Key Methodological Strength	Key Methodological Weakness	Overall RoB
Yin-Hao Lee et al. (2023) [4]	ECG	Both	Internal and external (multicentre) validation; integrates ECG with cardiovascular risk factors	Lowest reported AUC (0.72); high risk of bias in analysis domain	High
Leasure et al. (2021) [6]	ECG	Internal	Highest ECG AUC (0.961); strong balanced accuracy	Internal validation only; single-centre, limiting generalisability	High
Hyun-Gyu Lee et al. (2023) [16]	ECG	Internal	Ensemble integrating clinical, laboratory and ECG data; outperformed traditional risk scores	Internal validation only; analysis-domain weaknesses	High
Huang et al. (2022) [17]	ECG	Internal	Well-balanced sensitivity and specificity (≈0.90)	Internal validation only; high risk of bias in analysis domain	High
Park et al. (2023) [18]	ECG	Internal	Deep features tailored to stable angina improved on clinical risk factors	Internal validation only; modest accuracy (0.757)	High
Yeh et al. (2025) [11]	ECG	External	External validation; detects CAD in ECGs read as normal	Spectrum bias from young healthy controls (18–39) inflating metrics	High
Guo et al. (2024) [7]	CCTA	Internal	Fully automated CT-FFR; rapid processing (<4 min)	Internal validation only; no independent external cohort	High
Kigka et al. (2022) [14]	CCTA	Internal	Only prospective study; integrates imaging with biohumoral markers	Small sample (n = 187) relative to model complexity; overfitting risk	Unclear
Wahab Sait and Dutta (2023) [13]	CCTA	Internal	Highest performance (AUC 0.97/0.96)	Datasets curated and class-balanced with internal 70/30 train–test split, likely yielding optimistic estimates versus real-world prevalence	High
Yang et al. (2023) [19]	CCTA	External	External validation; large reduction in reporting time	Low specificity (0.596); analysis-domain weaknesses	High
Kim et al. (2025) [20]	CCTA	External	External multicentre validation for acute chest pain	Excluded non-diagnostic MPR images, potentially inflating AUC (0.871)	High

Abbreviations: ECG—electrocardiogram; CCTA—coronary computed tomography angiography; AUC—area under the receiver operating characteristic curve; CT-FFR—computed tomography fractional flow reserve; MPR—multiplanar reconstruction; RoB—risk of bias.

3.4.1. ECG-Based Models

Among the six ECG-based studies, a range of modelling approaches were used, from traditional machine learning to deep learning architectures. AUC values ranged from 0.72 to 0.961, with Sensitivity ranging from 0.75 to 0.93 and Specificity from 0.62 to 0.903. The best performing ECG model was reported by Leasure et al. [6], achieving AUC 0.961 with a 95% CI of 0.936–1.00, Sensitivity 0.930, and Specificity 0.856 using their ECGio architecture, though this was based on internal validation only. Another notable model was the InceptionV3 CNN used by Huang et al. [17], which achieved an AUC of 0.869 with well balanced Sensitivity (0.899) and Specificity (0.903). Hybrid approaches combining traditional machine learning with deep learning were common, as seen in the work of Yin-Hao Lee et al. [4] and Hyun-Gyu Lee et al. [16]. In terms of validation, four studies used internal validation only, while two studies [4,11] applied external validation, representing a minority of the ECG-based evidence.

3.4.2. CCTA-Based Models

The five CCTA-based studies showed greater consistency in their use of advanced deep learning architectures, particularly those designed for computer vision tasks. AUC values ranged from 0.77 to 0.97, with Sensitivity ranging from 0.84 to 0.990 and Specificity from 0.596 to 0.993. The best performing CCTA model was reported by Wahab Sait and Dutta [13], achieving an AUC of 0.97 on their first dataset and an AUC of 0.96 on

their second, with Sensitivity and Specificity exceeding 0.98 using a combined UNet++ and YOLOv7 architecture. This study relied on internal validation, with each dataset evaluated using a 70/30 train–test split, so its near perfect metrics should be read with corresponding caution. Kim et al. [20] achieved an AUC of 0.871 using YOLOv4 on curved multiplanar reconstruction images, while Kigka et al. [14] reported an AUC of 0.82 using gradient boosting and support vector machines with recursive feature elimination. Notably, only two of the five CCTA studies [19,20] used external validation, while Guo et al. [7], Kigka et al. [14], and Wahab Sait and Dutta [13] relied on internal validation only. External validation was therefore only marginally more common among the CCTA studies (two of five) than among the ECG-based studies (two of six), and does not by itself indicate greater methodological maturity.

3.4.3. Comparative Analysis

When comparing the two modalities, CCTA-based models showed a slightly higher top performance, with an AUC of 0.97, compared to ECG-based models with AUC scores of 0.961, although the ranges overlapped considerably. These figures should not be taken as a direct comparison, as the included studies vary considerably in their patient populations, CAD outcome definitions, reference standards, and validation strategies [5], and none evaluated both modalities within the same cohort. External validation was uncommon in both modalities, applied in 40% of CCTA studies [19,20] and 33% of ECG studies [4,11], which provides little basis for concluding that either body of literature is more methodologically rigorous; both groups also showed high risk of bias in the analysis domain. In terms of architecture, ECG models tended to use hybrid approaches combining traditional machine learning with deep learning, while CCTA models mainly used specialised computer vision architectures such as UNet++, YOLO variants, and ResNet. The best performing models in both modalities achieved high AUC values (0.961 for ECG and 0.97 for CCTA), though both relied on internal validation only. These findings should be interpreted in light of the high risk of bias identified in the evaluation component of most studies, which suggests that the reported performance metrics may be somewhat inflated and may not directly reflect real world clinical performance.

3.5. Diagnostic Accuracy

A comparison of diagnostic accuracy across the included studies showed notable differences between modalities, with CCTA-based models generally demonstrating stronger and more consistent performance compared to ECG-based approaches. Among the six ECG-based studies, ACC ranged from 0.69 for the hybrid model developed by Yin-Hao Lee et al. [4] to 0.957 for the multiclass prediction framework proposed by Leasure et al. [6]. Deep learning models using 2D image inputs showed solid performance; for example, Huang et al. [17] reported an ACC of 0.90, while Yeh et al. [11] achieved a validation ACC of 0.77. In contrast, the model focusing on stable angina by Park et al. [18] produced a more modest ACC (0.757) on the test set.

CCTA-based models consistently achieved high technical accuracy, frequently exceeding 0.80. The combined UNet++ and YOLOv7 framework developed by Wahab Sait and Dutta [13] reached a peak ACC of 0.995. Other automated frameworks also showed reliable performance, with Guo et al. [7] and Kigka et al. [14] reporting ACCs of 0.82 and 0.81, respectively. A full summary of these metrics alongside other performance variables is provided in Table 4. These accuracy values warrant careful interpretation, as the PROBAST+AI assessment rated 90.9% of the included studies as high risk of bias in the evaluation component, driven mainly by weaknesses in the analysis domain; the reported figures may

therefore be somewhat optimistic and unlikely to fully reflect performance achievable in routine clinical practice.

4. Discussion

4.1. Summary of Main Findings

Addressing RQ1, this systematic review of 11 primary studies shows that AI models using ECG and CCTA data achieve high diagnostic accuracy in predicting CAD, often outperforming traditional clinical risk models recommended by international guidelines such as the 2019 European Society of Cardiology (ESC) Guidelines [21]. However, these performance figures are promising but should not be taken as definitive, given that 90.9% of the included studies were rated as high risk of bias in the evaluation component under the PROBAST+AI assessment, largely because of weaknesses in the analysis domain; the reported accuracy is therefore likely to somewhat overstate what these models would achieve in real world practice. In the ECG-based domain, six studies used a range of architectures, including 1D and 2D convolutional neural networks (CNNs), XGBoost, and ensemble stacking, to analyse 12-lead waveforms or digitalised images. These models achieved AUCs ranging from 0.72 to 0.961, with Sensitivity between 0.75 and 0.93. Diagnostic accuracy further supports the potential of AI in cardiovascular screening. While ECG-based models serve as effective screening tools, with ACC reaching up to 0.957 [6], CCTA-based models demonstrated a higher ceiling of technical performance, peaking at ACC 0.995 [13]. These findings suggest that AI can reliably automate the identification of significant coronary lesions, often exceeding the baseline diagnostic yield of conventional clinical interpretation. Consistent with the findings reported by Siontis et al. [22], these algorithms detected subtle patterns in early depolarisation and repolarisation that are often not visible to the human eye. The best performing ECG model was the ECGio algorithm by Leasure et al. [6], which achieved an AUC of 0.961 on an internal validation set. While some ECG studies relied on internal data only, Yeh et al. [11] and Yin-Hao Lee et al. [4] additionally reported external validation. Yeh et al. [11] achieved an external validation AUC of 0.91, while Yin-Hao Lee et al. [4] reported an AUC of 0.72 on internal validation for their integrated ECG plus cardiovascular risk factor model, which was also evaluated on an independent external cohort, where it reached a lower AUC of 0.67. These findings, detailed in Tables 1, 2 and 4, highlight the potential of AI to improve pre-test probability assessments in outpatient settings.

For CCTA-based models, five studies focused on automated vessel extraction, stenosis grading, and functional assessment using computed tomography fractional flow reserve (CT-FFR). These models used advanced frameworks such as UNet++, YOLO variants (v4 and v7), and multi-task ResNet, achieving AUCs ranging from 0.77 to 0.97, with Sensitivity ranging from 0.84 to 0.990 and Specificity from 0.596 to 0.993. Wahab Sait and Dutta [13] achieved the strongest results using a YOLOv7 and optimised UNet++ framework, with AUCs of 0.97 and 0.96 across two internally validated datasets, each evaluated using a 70/30 train–test split. In line with the 2021 American Heart Association/American College of Cardiology (AHA/ACC) guidelines [23], which highlight the role of CCTA in rapid chest pain triage, these AI tools considerably reduced processing times, often to under four minutes, while maintaining a diagnostic accuracy comparable to expert readers.

A critical assessment of model quality using the PROBAST+AI framework (Figures 2–5) reveals notable systematic weaknesses despite the strong performance metrics reported. As defined by the PROBAST tool [24], model development was rated as having unclear concerns in 63.6% of cases, mainly due to insufficient reporting on calibration and data handling. The most concerning finding was in the evaluation domain, where 90.9% of studies were classified as high risk of bias. The Analysis domain was the main weakness,

as studies frequently failed to adequately handle missing data or relied on single-centre datasets with high pre-test probabilities. Therefore, while the performance metrics in Table 4 are promising, the high risk of bias identified through the PROBAST+AI framework suggests that these models require more standardised prospective validation before they can be widely adopted in clinical practice in line with current medical standards.

4.2. Interpretation of Findings

The findings of this review indicate a significant shift in the use of AI for CAD detection. While performance metrics for both modalities are high, external validation was limited in both, so differences in field maturity and validation rigour are less clear cut than the headline performance figures might imply; the modalities do, however, differ in their typical clinical role. As shown in Table 4, CCTA-based models generally achieve a higher and more consistent performance ceiling with AUC ranging from 0.77 to 0.97, compared to ECG-based models which have AUCs ranging from 0.72 to 0.961. This apparent gap is clinically plausible given that CCTA provides direct anatomic visualisation whereas ECG offers only an indirect physiological correlate of disease. That said, it cannot be treated as evidence of one modality's superiority, since a rigorous comparison would require both modalities to be evaluated in the same population using the same reference standard, outcome definition, and validation design [9].

4.2.1. Field Maturity and the Validation Gap

Turning to RQ2, external validation was uncommon in both modalities. Only 40% of CCTA studies [19,20] and 33% of ECG studies [4,11] evaluated their models on an independent external cohort, with the remaining studies relying on internal validation or single-centre holdout sets. This similarity, together with the high risk of bias in the analysis domain identified for 90.9% of studies, means the present evidence does not support the conclusion that either modality is closer to clinical readiness on the grounds of validation rigour. The tasks that CCTA-based models automate, such as vessel extraction and stenosis grading, are nonetheless already central to the radiological workflow described in the 2021 AHA/ACC guidelines [23], which highlight CCTA as a primary tool for rapid chest pain triage; this workflow fit may ease their eventual integration, but does not itself represent a demonstrated advantage in validation maturity.

In contrast, ECG-based AI serves as an electrical proxy for anatomical disease. As noted by Siontis et al. [22], AI can detect subtle repolarisation and depolarisation patterns that are not visible to the human eye, which explains why models such as that by Yeh et al. [11] can achieve an external validation AUC of 0.91, even in patients with a normal resting ECG. However, the AUC of 0.961 reported by Leasure et al. [6] should be interpreted with caution, as without external validation it remains a proof of concept rather than a generalisable clinical tool.

The difference in accuracy between the two modalities likely reflects the nature of the input data. CCTA provides direct anatomical evidence of plaque burden, allowing computer vision architectures such as YOLO and UNet++ to achieve near perfect accuracy in vessel segmentation and stenosis grading [7,13]. In contrast, the lower accuracy seen in some ECG models [4] highlights the difficulty of using electrophysiological proxies to predict anatomical disease in stable populations. However, as shown by Park et al. [18], the use of deep features specific to stable angina can considerably improve predictive accuracy over traditional clinical risk factors. It is also important to note that high accuracy figures reported in single-centre studies without external validation [6] may be inflated by the methodological weaknesses and high risk of bias identified in the PROBAST+AI assessment (Figure 5).

4.2.2. Direct vs. Proxy Data Interpretation

The performance difference between modalities is likely a result of the nature of the input data rather than algorithmic superiority. CCTA provides direct anatomical evidence of plaque burden and luminal narrowing, whereas ECG signals are indirect markers of ischaemia, making the prediction task more complex in stable populations where resting electrical activity may remain unchanged despite significant stenosis. The high Specificity values observed in CCTA studies, such as the 0.993 reported by Wahab Sait and Dutta [13], are consistent with the 2019 ESC guidelines [21], which value CCTA for its high NPV in ruling out disease.

4.2.3. The Role of Multimodal Integration

This review highlights that AI performance improves considerably when algorithms incorporate a broader clinical picture. Studies that combined signal or image data with clinical variables [4,14,16] achieved more robust results than those using the modality data alone. For example, Hyun-Gyu Lee et al. [16] found that adding the Atherogenic Index of Plasma (AIP) to ECG waveforms improved AUC to 0.767, outperforming traditional pre-test probability models such as the CAD Consortium score. This supports the ESC guideline recommendation to use risk modifiers such as diabetes and dyslipidaemia to refine pre-test probability assessments.

4.2.4. Methodological Rigour: PROBAST+AI and the Analysis Domain

Addressing RQ3, despite the high accuracy reported, the PROBAST+AI assessment (Figures 4 and 5) reveals that 90.9% of the included studies carry a high risk of bias in the evaluation domain. As defined by Wolff et al. [24], weaknesses in the Analysis domain, particularly the absence of calibration curves and inadequate handling of missing data, suggest that the reported performance figures are likely to be somewhat optimistic. In clinical terms, this means these models may not maintain their predictive accuracy when applied in lower prevalence populations. These reporting gaps highlight a broader need for transparency and standardised validation to ensure the clinical reproducibility of AI algorithms. Until these methodological limitations are addressed through prospective multicentre trials, these AI tools should be used as clinical decision support aids rather than standalone diagnostic replacements.

4.2.5. From Discrimination to Clinical Readiness

High discrimination, reflected in strong AUC values, does not by itself establish clinical readiness, and several further conditions need to be met before these models can be safely integrated into cardiovascular care. Calibration is perhaps the most pressing of these; a model may rank patients correctly yet still produce predicted probabilities that diverge from the true observed frequency of disease [24]. None of the included studies reported comprehensive calibration curves, which undermines confidence that their predicted probabilities would hold up in populations with a different disease prevalence, something the PROBAST+AI framework flags as a particular concern for AI-based models [25]. Explainability is another consideration that cannot be overlooked. Many of the stronger performing models relied on deep convolutional and multi-task architectures that function largely as opaque systems; the black-box nature of such models remains a genuine obstacle to clinician trust and wider clinical adoption [26], and without transparent reasoning or feature attribution, errors become harder to spot and audit. Performance is also tied to the quality of the underlying ECG signals and CCTA images, meaning accuracy achieved under favourable acquisition conditions may not carry over to routine practice, where signal and image quality are far more variable. On top of these technical issues, practical

barriers remain, including integration into existing clinical workflows and information systems, prospective validation within the intended care setting, and the regulatory and governance requirements that apply to clinical decision support tools. Taken together, the high discrimination reported across the included models is a necessary starting point for clinical adoption, but it is far from sufficient on its own.

4.3. Clinical Implications

The clinical application of AI models for CAD detection offers a tiered approach to cardiovascular care, with ECG-based models serving as low-cost screening tools in primary care and CCTA-based models providing rapid, high-resolution triage in acute and specialised settings. As shown in Table 4, the performance of these models suggests they are ready to support, though not yet replace, the clinical pathways established by the 2019 ESC and 2021 AHA/ACC guidelines [21,23].

4.3.1. Primary Care and Outpatient Risk Stratification

In the primary care setting, a key implication of this research is the potential to detect CAD signatures hidden within normal resting ECGs. Yeh et al. [11] showed that an AI algorithm could achieve an external validation AUC of 0.91 in patients whose ECGs were reported as normal by traditional standards, by identifying subtle repolarisation changes not visible to the human eye. The integration of clinical variables also considerably improves screening accuracy. Yin-Hao Lee et al. [4] demonstrated that combining 12-lead ECG data with cardiovascular risk factors (CVRFs) improved AUC to 0.72, while Hyun-Gyu Lee et al. [16] found that an ensemble model incorporating laboratory markers such as the Atherogenic Index of Plasma outperformed traditional risk scores including the Diamond–Forrester and CAD Consortium models. Consistent with the findings of Siontis et al. [22], these results suggest that AI-based ECG analysis can act as an effective screening tool, improving pre-test probability assessment and reducing unnecessary referrals for costly imaging in low to intermediate risk populations.

4.3.2. Emergency Department and Radiology Workflow Efficiency

For CCTA-based AI, the clinical implications centre on rapid triage and workflow improvement. In emergency department settings, Kim et al. [20] showed that a YOLOv4-based model achieved an external validation AUC of 0.871 for detecting obstructive CAD in patients presenting with acute chest pain, offering a reliable tool for rapid rule-out. In the radiology setting, automation helps address the growing demand for image analysis, as reflected in Table 5. Yang et al. [19] reported that AI reduced the average post-processing and reporting time by 76.32%, while Guo et al. [7] introduced fully automated CT-FFR technology that processed cases in under four minutes. These systems reduce inter-observer variability and are consistent with the 2023 ESC guidelines [27], which emphasise the value of CCTA when cardiac troponin results are inconclusive.

4.3.3. Barriers to Adoption

Despite the strong performance metrics shown in Table 4, several barriers to clinical adoption remain. External validation was limited across both modalities, so neither is yet close to routine deployment, and both still require large scale prospective evidence. More broadly, the high risk of bias identified across the included studies (Figures 4 and 5) means AI should currently be treated as a clinical decision support tool rather than a standalone diagnostic replacement, with clinicians remaining alert to potentially optimistic performance estimates and continuing to apply AI findings alongside clinical judgement and guideline-directed medical therapy [28].

4.4. Strengths and Limitations of the Review

The main strength of this rapid systematic review is its methodological rigour and adherence to the PRISMA-ScR guidelines [29], which ensures a transparent and reproducible synthesis of AI applications in CAD prediction. The use of a clear PICOS framework and a comprehensive search across three major databases (PubMed, Web of Science, and Scopus) allowed this review to capture the full scope of contemporary AI advances from 2021 to 2025. A key feature of this review is the application of the PROBAST+AI tool (Figures 2–5), which enabled a detailed assessment of both development quality and risk of bias, addressing the specific need for thorough quality appraisal of AI-based predictive models [25]. In addition, the structured data extraction into four distinct tables (Tables 1–4) supports the first direct comparative analysis of ECG-based AI models compared with CCTA-based models using a consistent synthesis framework. This contribution can be positioned against the existing review literature. Prior reviews have tended to address AI in cardiovascular diagnosis either across a broad range of imaging modalities or within a single modality. For example, Doolub et al. [9] provided a narrative review of AI across multiple non-invasive imaging modalities for CAD assessment, including CCTA, single-photon emission computed tomography (SPECT), and MRI, while Dey et al. [28] offered a state-of-the-art overview of artificial intelligence in cardiovascular imaging more generally, and Siontis et al. [22] focused specifically on AI-enhanced electrocardiography in cardiovascular disease management. These studies provide valuable overviews of the field; however, they are narrative or state-of-the-art in format rather than systematic. Notably, the review by Doolub et al. [9] did not apply a formal risk-of-bias appraisal to the studies it discussed, nor did it place ECG-based and CCTA-based models side by side. The present review is distinguished from these by two features in combination: a direct, structured comparison of ECG-based and CCTA-based AI models within a single PRISMA-guided synthesis, and a formal methodological appraisal of every included study using the PROBAST+AI tool. This combination allows the present work not only to summarise reported performance but also to weigh it against the risk of bias identified in each study, an emphasis that is largely absent from previous reviews in this area. Applicability concerns across the participants, predictors, and outcome domains were all rated as low, indicating that the findings from the 11 included studies are applicable to real-world clinical populations with suspected CAD.

However, several limitations should be acknowledged to ensure a balanced interpretation of these findings. First, the literature search relied primarily on title-based search terms. While this kept the review focused and rapid, it does reduce search sensitivity, and relevant studies may have been missed if they described the disease or methods using alternative terminology, such as coronary stenosis, ischaemia, obstructive coronary disease, coronary CTA, CT-FFR, automated detection, segmentation, or deep neural network, or if the relevant AI-related terms appeared only in the abstract or keywords rather than the title. This may have affected how completely studies were retrieved, though searching across three major databases and using deliberately broad population and intervention terms was intended to partly offset this. Second, study selection, data extraction, and quality assessment were all carried out by a single reviewer, which, despite the use of a systematic protocol, may introduce bias compared with the dual independent reviewer approach recommended by standard methodological guidelines. To reduce this risk, all stages were guided by the predefined internal protocol and PICOS eligibility criteria, uncertain cases were set aside and revisited in a dedicated session against these fixed criteria, and a detailed exclusion log was kept at the full-text stage. That said, the single-reviewer process may still have affected study retrieval, inclusion decisions, data extraction, and the PROBAST+AI risk-of-bias assessment; a second independent reviewer could have identified additional eligible studies or reached different judgements at any of these stages, and this possibility

should be kept in mind when interpreting the findings. Third, the restriction to English-only articles and the exclusion of conference abstracts may have led to the omission of relevant non-English research or emerging work not yet published as full peer reviewed articles. Fourth, the narrow date range from 2021 to 2025 was intentionally chosen to focus on the most recent models, but this means earlier foundational work in the field was not included, which may limit the historical context of algorithm development. Furthermore, the considerable variation in model architectures, ranging from 1D CNNs for waveform analysis [16] to YOLOv7 for 3D image processing [13], made it unsuitable to conduct a meta-analysis, limiting this review to a narrative synthesis without pooled effect estimates [30]. Finally, by focusing solely on ECG and CCTA, this review does not cover AI advances in other important diagnostic modalities such as echocardiography, magnetic resonance imaging (MRI), or SPECT, which were noted as alternative screening tools in the studies by Leasure et al. [6] and Yeh et al. [11]. Despite these limitations, the transparency of the reporting ensures this review remains a reliable evidence base for evaluating current AI performance in CAD detection.

4.5. Limitations of the Included Evidence

The synthesis of evidence in this review is constrained by significant methodological weaknesses identified through the PROBAST+AI assessment (Figures 2–5). While Table 4 reports high diagnostic performance, the overall quality of model development and evaluation points to a high risk-of-optimism bias, where reported accuracy values likely exceed real-world performance due to identified shortcomings in calibration and data handling [24]. As shown in Figure 5, 90.9% of the included studies were rated as high risk of bias in the evaluation domain, primarily because none of the included studies achieved a low risk rating in the Analysis domain.

A further limitation concerns validation rigour, which was limited across both modalities rather than confined to one. As noted in Table 3, 67% of ECG-based studies relied solely on internal validation or single-centre holdout sets [6,16], and external validation was similarly uncommon among CCTA-based studies, applied in only 40% of them, with the remainder [7,13,14] likewise relying on internal validation; as outlined by Wolff et al. [24], internal validation alone is insufficient to establish clinical generalisability. Most CCTA studies also remained retrospective in design. Kigka et al. [14] provided the only prospective evidence; however, their findings are limited by a small sample size of 187 patients relative to the complexity of the gradient-boosting framework used, which increases the risk of overfitting.

The evidence is also affected by spectrum bias and clean data bias. Yeh et al. [11] used a group of young healthy controls aged 18 to 39 to balance their dataset, a design choice that can artificially inflate performance metrics compared to a real-world symptomatic cohort in which the non-CAD group often presents with significant comorbidities [25]. Similarly, Kim et al. [20] excluded non-diagnostic curved MPR images from their evaluation. This is clinically problematic, because real-world emergency department scans are frequently affected by the very artefacts, such as motion and calcification, that were excluded, which means the reported AUC (0.871) may not be representative of routine clinical practice.

A further limitation specific to the ECG-based evidence concerns signal quality and generalisability. The diagnostic value of an ECG is highly dependent on acquisition conditions, and technical factors such as electrode misplacement, motion artifact, and improper filter application are common in routine practice and can meaningfully alter the recorded signal [31]. These effects are not uniform across patients; accurate precordial lead placement is frequently complicated by body habitus (BMI), sex, age, and race/ethnicity, introducing systematic variation linked to these patient characteristics [31]. This variability is made

worse by the absence of a universally reliable standard for automated ECG interpretation. In one recent evaluation, roughly 39% of ECGs were interpreted incorrectly by automated software, with errors varying by device type and increasing in older patients [32]. Most of the included ECG studies were developed and tested on curated datasets, so their reported performance may not fully account for this kind of real-world signal variability. Models trained on selected high-risk cohorts may also perform less well in broader populations; studies have shown that deriving models from cohorts enriched for high-risk patients while excluding low- and moderate-risk cases risks inflating accuracy and limiting real-world applicability [33]. All of this reinforces the need for careful external validation of ECG-based models across diverse acquisition settings and patient groups before any move toward clinical deployment.

The CCTA-based evidence carries its own limitations rooted in image quality and case selection. CCTA is inherently sensitive to acquisition and patient factors, and image artefacts are common even in routine practice; in one dual-source CCTA series, artefacts of some kind were present in 41.3% of studies, with calcified plaque the single leading source, and their occurrence was significantly associated with coronary calcium score, change in heart rate during scans, patient age, and body mass index [34]. Coronary calcification is a particularly important constraint. Blooming artefact progressively erodes diagnostic performance as calcium burden rises, and reported specificity has been shown to fall from 99.8% in minimally calcified plaques to 67.9% in the most heavily calcified, to the point where CCTA may need to be supplemented with calcium scoring or invasive angiography in such cases [35]. These vulnerabilities matter when interpreting the included models, because performance obtained on high-quality curated scans may not transfer to unselected clinical populations, where motion, arrhythmia, suboptimal contrast timing, and non-diagnostic image quality are far more common. This risk is made worse by case selection during evaluation; several studies excluded challenging images from analysis, whether non-diagnostic curved multiplanar reconstructions in the case of Kim et al. [20] or artefact-affected segments removed for quality reasons in diagnostic accuracy work more generally [35]. Since such exclusions remove precisely the borderline cases that are hardest to classify, they can inflate apparent accuracy and produce estimates that overstate real world performance. All of this reinforces the need for prospective validation on consecutive, unselected CCTA data acquired under routine conditions before these models can be relied upon for clinical decision making.

Furthermore, the absence of standardised reporting for model calibration and confidence intervals remains a widespread problem across the included studies. None of the 11 studies provided comprehensive calibration curves, which are essential for determining whether a model's predicted probability aligns with the true observed frequency of disease [25]. In addition, the variation in outcome definitions, ranging from $\geq$50% diameter stenosis [4] to >70% [17], prevents a direct meta-analysis. These limitations, summarised in the risk of bias figures (Figures 4 and 5), indicate that while AI models for CAD show strong technical promise, the current evidence base lacks the prospective, multicentre, and calibrated validation required for inclusion in clinical guidelines such as the 2019 ESC standards [21].

Finally, this lack of comparability points to a deeper structural gap in the field: the absence of standardised, publicly available benchmark datasets with consensus reference annotations for AI-based CAD prediction. The value of such standardisation is well illustrated in neuro-oncological imaging, and the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) was created precisely because competing algorithms had been validated on small private datasets using inconsistent metrics and differing imaging inputs, making objective comparison between methods highly challenging [36]. By providing a

common dataset and a unified evaluation framework, benchmarks of this kind allow different models to be compared directly on the same data, and they also expose the intrinsic difficulty of the task; in BRATS, even expert human raters disagreed considerably when annotating tumour sub-regions, with inter-rater Dice scores in the range of 74–85% [36]. The AI-based CAD prediction field currently lacks an equivalent shared benchmark for either ECG or CCTA data, and this is a central reason why the cross-study performance differences summarised in Table 4 cannot be read as head-to-head comparisons. Developing standardised, multi-institutional datasets with harmonised CAD outcome definitions and consensus labelling would therefore be a meaningful step toward enabling fair, reproducible comparisons between models and, in turn, more reliable evidence synthesis.

4.6. Future Research Directions

To move AI-based CAD models from experimental tools to standard clinical practice, future research must prioritise large-scale prospective multicentre trials that address the widespread high risk of bias identified in the Analysis domain (Figures 3 and 5). While a slightly higher proportion of CCTA-based models used external validation than ECG-based models, external validation remained uncommon in both, and most studies were retrospective and prone to spectrum bias (Table 1). Given that external validation was used in only 33% of ECG-based studies and 40% of CCTA-based studies, dedicated multicentre efforts are urgently needed across both modalities to validate promising algorithms, such as those by Yeh et al. [11] and Leasure et al. [6], in diverse and unselected populations before widespread clinical use. Methodological transparency should be improved by consistently applying the PROBAST+AI framework [25], particularly through the routine reporting of calibration curves to ensure that predicted probabilities reflect the true observed frequency of disease. Researchers should also move away from clean data designs, such as the artefact exclusion seen in the work of Kim et al. [20], and develop more robust models capable of handling real-world clinical noise and diverse multiethnic populations, as recommended by the 2019 ESC Guidelines [21].

A significant and largely unexplored opportunity lies in developing multimodal fusion AI that combines physiological signals from ECG with anatomical information from CCTA into a single diagnostic pathway. Building on the work of Hyun-Gyu Lee et al. [16] and Kigka et al. [14], who demonstrated the value of integrating modality specific data with biohumoral and clinical markers, future studies should examine whether a unified AI model can serve as a complete diagnostic tool in line with the precision medicine goals of the 2021 AHA/ACC guidelines [23]. Finally, the field must engage with implementation science to assess the impact of AI on clinician decision making, patient length of stay in emergency departments as suggested by Kim et al. [20], and overall cost effectiveness [28]. Prospective trials that go beyond technical accuracy to measure real clinical outcomes will be essential to support the deployment of these tools within modern healthcare systems.

Finally, the methodology of systematic reviews in this field is itself likely to evolve. Because the present review was conducted rapidly and relied on single-reviewer screening, it illustrates the scalability pressures that increasingly motivate the use of large language model (LLM) and other AI-assisted screening workflows. Early evaluations suggest such tools can considerably speed up the screening process while retaining acceptable accuracy; in one prospective diagnostic study, LLM-assisted title and abstract screening reduced the processing time for 100 records from 17.2 to 1.3 min, and refining the screening prompt raised sensitivity from 0.75 to 0.91 [37]. In this field, such tools could improve both the scalability and sensitivity of study identification, for example by handling larger volumes of records or recovering studies that use varied terminology and would be missed by a narrow title-based search. That said, the same evidence makes clear that responsible

use depends on preserving the methodological safeguards that underpin conventional review conduct: transparent and pre-specified eligibility criteria, reproducible and clearly documented prompts or procedures, complete audit trails of automated decisions, and human adjudication for final inclusion, since automated screening remains imperfect and prompt design materially affects performance [37]. Within these constraints, AI-assisted screening could help future reviews in AI-based CAD prediction achieve broader and more reproducible coverage than was feasible here, while maintaining the rigour that evidence synthesis requires.

5. Conclusions

This rapid systematic review assessed the diagnostic performance, comparative value, and methodological quality of AI models developed for CAD prediction using ECG and CCTA data. Both modalities showed strong diagnostic accuracy, with ECG-based models achieving AUCs ranging from 0.72 to 0.961 and CCTA-based models achieving AUCs ranging from 0.77 to 0.97. External validation was uncommon in both groups, being used in 40% of CCTA studies and 33% of ECG-based studies, so the evidence does not establish that either modality has greater validation maturity. However, the PROBAST+AI assessment revealed a high risk of bias in 90.9% of all included studies, mainly driven by weaknesses in the analysis domain, including poor handling of missing data and the absence of model calibration reporting. These findings suggest that AI models should currently be used as clinical decision support tools rather than standalone diagnostic replacements. ECG-based AI is best suited as a low-cost screening tool in primary care, while CCTA-based AI offers value for rapid triage in acute settings. Future research should focus on prospective multicentre validation and the development of multimodal approaches that combine physiological and anatomical data to improve diagnostic precision and clinical applicability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/aimed-4307961/s1, Table S1: PRISMA Checklist.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created in this study. The data extracted from included studies are available within the article, in Tables 1–5, or from the corresponding author upon reasonable request. This review was prospectively registered on the Open Science Framework, https://doi.org/10.17605/OSF.IO/H793Y.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Abbreviations

AHA/ACC: American Heart Association/American College of Cardiology; AI: Artificial Intelligence; AIP: Atherogenic Index of Plasma; AUC: Area Under the Curve; CAD: Coronary Artery Disease; CAG: Invasive Coronary Angiography; CCTA: Coronary Computed Tomography Angiography; CI: Confidence Interval; CNN: Convolutional Neural Network; CT-FFR: Computed Tomography Fractional Flow Reserve; CVRFs: Cardiovascular Risk Factors; DL: Deep Learning; ECG: Electrocardiogram; ED: Emergency Department; EN: Elastic Net; ESC: European Society of Cardiology; LDL: Low-Density Lipoprotein; LR: Logistic Regression; ML: Machine Learning; MPR: Multiplanar Reconstruction; MRI: Magnetic Resonance Imaging; NPV: Negative Predictive Value; NT-proBNP: N-terminal pro-B-type Natriuretic Peptide; PICOS: Population, Intervention, Comparison, Outcome,

Study Design; PROBAST+AI: Prediction model Risk Of Bias ASsessment Tool updated for Artificial Intelligence methods; PTP: Pre-Test Probability; ResNet: Residual Network; ROC: Receiver Operating Characteristic; SPECT: Single Photon Emission Computed Tomography; SVM: Support Vector Machine; TSH: Thyroid-Stimulating Hormone;

References

1. Taha, K.; Ross, H.J.; Peikari, M.; Mueller, B.; Fan, C.P.S.; Crowdy, E.; Moayedi, Y.; Billia, F.; Manlhiot, C. Predicting the future risk and outcomes of severe heart failure and coronary artery disease with machine learning in the UK Biobank cohort. *PLoS ONE* **2025**, *20*, e0329461. [CrossRef]
2. Muhammad, L.J.; Al-Shourbaji, I.; Haruna, A.A.; Mohammed, I.A.; Ahmad, A.; Jibrin, M.B. Machine learning predictive models for coronary artery disease. *SN Comput. Sci.* **2021**, *2*, 350. [CrossRef] [PubMed]
3. Shih, K.; Herz, N.; Sheikh, A.; O'Neill, C.; Carter, P.; Anderson, M. Economic burden of cardiovascular disease in the United Kingdom. *Eur. Heart J. Qual. Care Clin. Outcomes* **2025**, *11*, 678–690. [CrossRef] [PubMed]
4. Lee, Y.H.; Hsieh, M.T.; Chang, C.C.; Tsai, Y.L.; Chou, R.H.; Lu, H.H.S.; Huang, P.H. Improving detection of obstructive coronary artery disease with an artificial intelligence-enabled electrocardiogram algorithm. *Atherosclerosis* **2023**, *381*, 117238. [CrossRef] [PubMed]
5. Yu, Y.; Li, W.; Wu, J.; Hua, X.; Jin, B.; Shi, H.; Chen, Q.; Pan, J. Machine learning models using symptoms and clinical variables to predict coronary artery disease on coronary angiography. *Adv. Interv. Cardiol.* **2024**, *20*, 30–36. [CrossRef] [PubMed]
6. Leasure, M.; Jain, U.; Butchy, A.; Otten, J.; Covalesky, V.A.; McCormick, D.; Mintz, G.S. Deep learning algorithm predicts angiographic coronary artery disease in stable patients using only a standard 12-lead electrocardiogram. *Can. J. Cardiol.* **2021**, *37*, 1715–1724. [CrossRef] [PubMed]
7. Guo, B.; Jiang, M.; Guo, X.; Tang, C.; Zhong, J.; Lu, M.; Liu, C.; Zhang, X.; Qiao, H.; Zhou, F.; et al. Diagnostic and prognostic performance of artificial intelligence-based fully-automated on-site CT-FFR in patients with CAD. *Sci. Bull.* **2024**, *69*, 1472–1485. [CrossRef] [PubMed]
8. Yao, S.; Zang, J.; Hao, Q.; Wang, J.; Zhang, Z.; Xue, C. Machine learning-based CAD detection using integrated ECG and PCG parameter features. *BioMed Phys. Eng. Express* **2025**, *11*, 055026. [CrossRef] [PubMed]
9. Doolub, G.; Mamalakis, M.; Alabed, S.; Van der Geest, R.J.; Swift, A.J.; Rodrigues, J.C.; Garg, P.; Joshi, N.V.; Dastidar, A. Artificial intelligence as a diagnostic tool in non-invasive imaging in the assessment of coronary artery disease. *Med. Sci.* **2023**, *11*, 20. [CrossRef] [PubMed]
10. Hajianfar, G.; Gharibi, O.; Sabouri, M.; Mohebi, M.; Amini, M.; Yasemi, M.J.; Chehreghani, M.; Maghsudi, M.; Mansouri, Z.; Edalat-Javid, M.; et al. Artificial intelligence-powered coronary artery disease diagnosis from SPECT myocardial perfusion imaging: A comprehensive deep learning study. *Eur. J. Nucl. Med. Mol. Imaging* **2025**, *52*, 3019–3035. [CrossRef] [PubMed]
11. Yeh, C.H.; Tsai, T.H.; Chen, C.H.; Chou, Y.J.; Mao, C.T.; Su, T.P.; Yang, N.I.; Lai, C.C.; Chen, C.T.; Sytwu, H.K.; et al. Artificial intelligence-enhanced electrocardiography improves the detection of coronary artery disease. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 278–286. [CrossRef] [PubMed]
12. Park, J.; Kim, J.; Kang, S.H.; Lee, J.; Hong, Y.; Chang, H.J.; Cho, Y.; Yoon, Y.E. Artificial intelligence-enhanced electrocardiography analysis as a promising tool for predicting obstructive coronary artery disease in patients with stable angina. *Eur. Heart J. Digit Health* **2024**, *5*, 444–453. [CrossRef] [PubMed]
13. Wahab Sait, A.R.; Dutta, A.K. Developing a deep-learning-based coronary artery disease detection technique using computer tomography images. *Diagnostics* **2023**, *13*, 1312. [CrossRef] [PubMed]
14. Kigka, V.I.; Georga, E.; Tsakanikas, V.; Kyriakidis, S.; Tsompou, P.; Siogkas, P.; Michalis, L.K.; Naka, K.K.; Neglia, D.; Rocchiccioli, S.; et al. Machine learning coronary artery disease prediction based on imaging and non-imaging data. *Diagnostics* **2022**, *12*, 1466. [CrossRef] [PubMed]
15. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan: A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [CrossRef] [PubMed]
16. Lee, H.G.; Park, S.D.; Bae, J.W.; Moon, S.; Jung, C.Y.; Kim, M.S.; Kim, T.H.; Lee, W.K. Machine learning approaches that use clinical, laboratory, and electrocardiogram data enhance the prediction of obstructive coronary artery disease. *Sci. Rep.* **2023**, *13*, 12635. [CrossRef] [PubMed]
17. Huang, P.S.; Tseng, Y.H.; Tsai, C.F.; Chen, J.J.; Yang, S.C.; Chiu, F.C.; Chen, Z.-W.; Hwang, J.-J.; Chuang, E.Y.; Wang, Y.-C.; et al. An artificial intelligence-enabled ECG algorithm for the prediction and localization of angiography-proven coronary artery disease. *Biomedicines* **2022**, *10*, 394. [CrossRef] [PubMed]
18. Park, J.; Yoon, Y.; Cho, Y.; Kim, J. Feasibility of artificial intelligence-based electrocardiography analysis for the prediction of obstructive coronary artery disease in patients with stable angina: Validation study. *JMIR Cardio* **2023**, *7*, e44791. [CrossRef] [PubMed]

19. Yang, W.; Chen, C.; Yang, Y.; Chen, L.; Yang, C.; Gong, L.; Wang, J.; Shi, F.; Wu, D.; Yan, F. Diagnostic performance of deep learning-based vessel extraction and stenosis detection on coronary computed tomography angiography for coronary artery disease: A multi-reader multi-case study. *Radiol. Med.* **2023**, *128*, 307–315. [CrossRef] [PubMed]
20. Kim, J.Y.; Park, J.; Lee, K.H.; Lee, J.W.; Park, J.; Kim, P.K.; Han, K.; Baek, S.-E.; Im, D.J.; Choi, B.W.; et al. Development and validation of deep learning model for detection of obstructive coronary artery disease in patients with acute chest pain: A multi-center study. *Radiol. Med.* **2025**, *130*, 1615–1624. [CrossRef] [PubMed]
21. Knuuti, J.; Wijns, W.; Saraste, A.; Capodanno, D.; Barbato, E.; Funck-Brentano, C.; Prescott, E.; Storey, R.F.; Deaton, C.; Cuisset, T.; et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur. Heart J.* **2020**, *41*, 407–477. [CrossRef] [PubMed]
22. Siontis, K.C.; Noseworthy, P.A.; Attia, Z.I.; Friedman, P.A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat. Rev. Cardiol.* **2021**, *18*, 465–478. [CrossRef] [PubMed]
23. Gulati, M.; Levy, P.D.; Mukherjee, D.; Amsterdam, E.; Bhatt, D.L.; Birtcher, K.K.; Blankstein, R.; Boyd, J.; Bullock-Palmer, R.P.; Conejo, T.; et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain. *J. Am. Coll. Cardiol.* **2021**, *78*, e187–e285. [CrossRef] [PubMed]
24. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; for the PROBAST Group. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann. Intern. Med.* **2019**, *170*, W1–W33. [CrossRef] [PubMed]
25. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Navarro, C.A.; Dhiman, P.; Beam, A.L.; Van Calster, B.; Celi, L.A.; Denaxas, S.; et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **2025**, *388*, e082505. [CrossRef] [PubMed]
26. Ayano, Y.M.; Schwenker, F.; Dufera, B.D.; Debelee, T.G. Interpretable machine learning techniques in ECG-based heart disease classification: A systematic review. *Diagnostics* **2022**, *13*, 111. [CrossRef] [PubMed]
27. Byrne, R.A.; Rossello, X.; Coughlan, J.J.; Barbato, E.; Berry, C.; Chieffo, A.; Claeys, M.J.; Dan, G.-A.; Dweck, M.R.; Galbraith, M.; et al. 2023 ESC guidelines for the management of acute coronary syndromes. *Eur. Heart J.* **2023**, *44*, 3720–3826. [CrossRef] [PubMed]
28. Dey, D.; Slomka, P.J.; Leeson, P.; Comaniciu, D.; Shrestha, S.; Sengupta, P.P.; Marwick, T.H. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **2019**, *73*, 1317–1335. [CrossRef] [PubMed]
29. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
30. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
31. García-Niebla, J.; Llontop-García, P.; Valle-Racero, J.I.; Serra-Autonell, G.; Batchvarov, V.N.; De Luna, A.B. Technical mistakes during the acquisition of the electrocardiogram. *Ann. Noninvasive Electrocardiol.* **2009**, *14*, 389–403. [CrossRef] [PubMed]
32. Kraik, K.; Dykiert, I.A.; Niewiadomska, J.; Ziemer-Szymańska, M.; Mikołajczak, K.; Kreń, M.; Kukiełka, P.; Martuszewski, A.; Harych, T.; Poręba, R.; et al. The most common errors in automatic ECG interpretation. *Front. Physiol.* **2025**, *16*, 1590170. [CrossRef] [PubMed]
33. Chang, C.H.; Lin, C.S.; Lee, C.H.; Lin, C.; Lee, C.C.; Liu, W.T.; Tsai, D.J. Real-world application of deep learning for ECG-based prediction of coronary artery disease and revascularization needs. *Eur. Heart J. Digit Health* **2025**, *6*, 1124–1133. [CrossRef] [PubMed]
34. Dey, D.; Lee, C.J.; Ohba, M.; Gutstein, A.; Slomka, P.J.; Cheng, V.; Suzuki, Y.; Suzuki, S.; Wolak, A.; Le Meunier, L.; et al. Image quality and artifacts in coronary CT angiography with dual-source CT: Initial clinical experience. *J. Cardiovasc. Comput. Tomogr.* **2008**, *2*, 105–114. [CrossRef] [PubMed]
35. Qi, L.; Tang, L.J.; Xu, Y.; Zhu, X.M.; Zhang, Y.D.; Shi, H.B.; Yu, R.-B. The diagnostic performance of coronary CT angiography for the assessment of coronary stenosis in calcified plaque. *PLoS ONE* **2016**, *11*, e0154852. [CrossRef] [PubMed]
36. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef] [PubMed]
37. Oami, T.; Okada, Y.; Nakada, T.A. Performance of a large language model in screening citations. *JAMA Netw. Open* **2024**, *7*, e2420496. [CrossRef] [PubMed]