Eurasian J. Chem. Med. Pet. Res., **(2027),** Volume 6, Issue 2, 241-254

Eurasian Journal of Chemical, Medicinal and Petroleum Research

Journal homepage: https://www.ejcmpr.com/

DOI: https://doi.org/10.5281/zenodo.24632201

Artificial Intelligence in Cardiovascular Disease Diagnosis: Predictive Biomarkers, Early Detection, and Personalized Clinical Risk Assessment

Reyhane Maleki
MSc in Medical-Surgical Nursing, Ilam University of Medical Sciences, Ilam, Iran

Article info
Received: 11.09.2026
Accepted: 26.11.2026
Available Online: 27.11.2026
Checked for Plagiarism: Yes

Keywords:
Artificial intelligence; cardiovascular disease; predictive biomarkers; early detection; risk assessment

ABSTRACT

Cardiovascular disease (CVD) remains the leading cause of global mortality, yet conventional diagnostic and risk-stratification paradigms frequently fail to identify high-risk individuals before adverse events occur. Artificial intelligence (AI) has emerged as a transformative technology capable of extracting latent diagnostic and prognostic information from routine clinical data, imaging, and multi-omics profiles. This review synthesizes current evidence on AI applications across three interconnected domains: predictive biomarker discovery, early disease detection, and personalized clinical risk assessment. Deep learning models applied to electrocardiography can identify subclinical ventricular dysfunction, atrial fibrillation, and structural heart disease with performance exceeding conventional interpretation. In cardiac imaging, AI enables automated quantification, plaque characterization, and detection of prognostically significant features invisible to human readers. Multi-omics integration through frameworks such as CardiOmicScore demonstrates that proteomic and metabolomic risk scores substantially enhance predictive accuracy up to 15 years before disease onset. Machine learning models incorporating clinical, lifestyle, and genetic data achieve area under the curve values of 0.85 for composite CVD prediction, outperforming the Framingham Risk Score by a clinically meaningful margin. Despite these advances, critical barriers persist: the predominance of retrospective single-center studies, limited external validation, demographic bias in training datasets, and a paucity of prospective implementation trials demonstrating outcome improvement. This review critically evaluates the current landscape, identifies methodological limitations, and proposes a translational framework prioritizing multicenter validation, explainability, and workflow integration to realize AI's potential for precision cardiovascular prevention.

Introduction

Cardiovascular disease (CVD) continues to impose an extraordinary global health burden, accounting for approximately 17.9 million deaths annually and representing the leading cause of mortality worldwide. Despite decades of therapeutic innovation and the establishment of evidence-based preventive strategies, a substantial proportion of cardiovascular events occur in individuals classified as low or intermediate risk by conventional assessment tools.

This paradox underscores a fundamental limitation of current diagnostic and prognostic paradigms: they rely on a restricted set of clinical variables, imaging findings, and laboratory biomarkers that capture only a fraction of the pathophysiological complexity underlying atherogenesis, arrhythmogenesis, and cardiac remodeling. The conventional diagnostic pathway for CVD typically begins with symptom assessment, electrocardiography (ECG), and echocardiography, followed by more advanced imaging or invasive testing when indicated.

Risk stratification tools such as the Framingham Risk Score and the Systematic Coronary Risk Evaluation (SCORE) model have served as cornerstones of primary prevention for decades, yet their predictive discrimination remains modest, with area under the curve (AUC) values typically ranging

Corresponding Author:* ***Reyhane Maleki* *(reyhane76.maleki@gmail.com)*

from 0.65 to 0.75. These instruments were developed using classical statistical methods that assume linear relationships among a limited number of predictors, an assumption that poorly reflects the multifactorial and interactive nature of cardiovascular pathophysiology. Furthermore, they fail to incorporate emerging biomarkers, genetic information, and digital data streams that may carry substantial prognostic signal.

Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), offers a fundamentally different approach to cardiovascular diagnosis and risk prediction. Rather than relying on predefined variable interactions, AI models can learn complex, high-dimensional patterns directly from data, potentially capturing subtle signals that elude conventional analysis. This capability is particularly relevant in cardiology, where the volume and diversity of available data from high-resolution imaging to continuous physiological monitoring to multi-omics profiling exceed the capacity of traditional analytical frameworks.

Three interconnected domains have emerged as particularly promising for AI application.

First, predictive biomarker discovery leverages AI to identify novel molecular signatures from metabolomic, proteomic, and genomic data that improve risk discrimination beyond established clinical markers. The Cardi Omic Score framework, for example, demonstrated that deep learning-derived proteomic and metabolomic risk scores achieved C-indices of 0.69-0.82 and 0.64-0.74 respectively across six common CVDs, with additive predictive value extending up to 15 years before clinical manifestation.

Second, early detection encompasses AI applications that identify subclinical disease before symptoms prompt clinical presentation. Deep learning models applied to standard 12-lead ECGs can detect latent left ventricular systolic dysfunction, hypertrophic cardiomyopathy, and paroxysmal atrial fibrillation, conditions that frequently escape conventional diagnostic scrutiny until advanced stages. Wearable devices augmented by AI algorithms enable continuous rhythm monitoring and have received regulatory clearance for atrial fibrillation screening in consumer populations.

Third, personalized clinical risk assessment integrates multimodal data clinical, lifestyle, genetic, and imaging to generate individualized risk estimates that outperform population-based algorithms. A recent UK Biobank analysis demonstrated that an integrated ML model combining these data layers achieved an AUC of 0.85 for composite CVD prediction, representing a 0.12 improvement over the Framingham Risk Score.

Despite this promise, the translation of AI from computational proof-of-concept to bedside clinical tool has been slower than enthusiasm might predict. A systematic review of AI applications in cardiovascular imaging reported that only 18.9% of studies made source code publicly available, and only 32.3% utilized exclusively public datasets or shared their data. The predominance of retrospective, single-center study designs limits generalizability, and concerns regarding algorithmic bias, demographic representativeness, and the "black-box" nature of deep learning models have prompted calls for greater transparency and explainability. Moreover, the evidence base for outcome improvement as opposed to diagnostic accuracy alone remains sparse, with few randomized controlled trials demonstrating that AI-guided care improves patient outcomes compared with standard management.

This review provides a critical synthesis of the current evidence on AI in cardiovascular disease diagnosis, with particular focus on predictive biomarkers, early detection strategies, and personalized risk assessment. We examine the methodological strengths and limitations of existing studies, evaluate the readiness of AI tools for clinical deployment, and propose a translational framework that prioritizes prospective validation, equity, and workflow integration. The overarching premise is that AI's value in cardiovascular medicine will be realized not through algorithmic sophistication alone, but through rigorous evidence generation that demonstrates meaningful improvements in diagnostic yield, risk discrimination, and ultimately patient outcomes.

Literature Review

The application of artificial intelligence to cardiovascular diagnosis and risk prediction has expanded rapidly over the past decade, encompassing diverse methodological approaches and clinical targets. This review synthesizes the evidence across three principal domains: predictive biomarker discovery, early detection through physiological and imaging data, and personalized risk assessment integrating multimodal information.

Predictive Biomarkers and Multi-omics Integration:

The search for biomarkers that improve upon established clinical predictors has been a central pursuit in preventive cardiology. Conventional biomarkers such as high-sensitivity C-reactive protein, B-type natriuretic peptide, and troponin have demonstrated modest incremental value, with most studies reporting C-index improvements of less than 0.05 when added to standard risk models. The emergence of high-throughput omics technologies

metabolomics, proteomics, and genomics has expanded the biomarker discovery space exponentially, but the analytical complexity of these high-dimensional datasets has simultaneously created a need for sophisticated computational methods.

Luo et al. addressed this challenge by developing Cardi Omic Score, a multitask deep learning framework trained on UK Biobank data comprising 168 metabolites and 2,920 proteins. The framework generated disease-specific risk scores (Met Score and Pro Score) for six common cardiovascular conditions: coronary artery disease, stroke, heart failure, atrial fibrillation, peripheral artery disease, and venous thromboembolism. The authors reported that Pro Score and Met Score were strong sole predictors, with C-indices ranging from 0.69 to 0.82 for Pro Score and 0.64 to 0.74 for Met Score. Critically, when combined with clinical data, multi-omics profiling significantly enhanced risk prediction across all CVDs, increasing the C-index by 0.005 to 0.102 up to 15 years before disease onset. This temporal window is particularly relevant for preventive interventions, as it suggests that AI-enhanced biomarker profiling can identify at-risk individuals well before the emergence of conventional risk factors or clinical symptoms.

The Cardi Omic Score study also identified specific proteins and metabolites with putative causal relevance to CVD pathogenesis, including several that had not previously been implicated in cardiovascular disease. These findings represent a form of "data-driven pathway discovery" wherein the model's predictive architecture highlights biological features that may warrant further mechanistic investigation. However, the authors appropriately noted that these findings require external validation in independent cohorts and that the incremental predictive value, while statistically significant, remains modest in absolute terms. The clinical utility of multi-omics profiling will depend not only on predictive performance but also on cost, accessibility, and the availability of actionable interventions for individuals identified as high-risk through omics-based assessment.

Early Detection Through Electrocardiography and Wearable Devices:

The electrocardiogram is among the most widely performed diagnostic tests in medicine, yet conventional interpretation captures only a fraction of the information embedded in the cardiac electrical signal. Deep learning models have demonstrated the capacity to extract latent diagnostic signatures from standard 12-lead ECGs, enabling detection of conditions that are not apparent to human readers. Garcia et al. reviewed the evidence for AI-ECG applications in preventive cardiology and reported that deep learning models can identify ventricular dysfunction, structural heart disease, atrial fibrillation, cardiomyopathy, and heart failure risk from sinus rhythm ECGs. Importantly, the authors distinguished between detection (identifying prevalent disease), prediction (forecasting future events), and prognostication (estimating risk of adverse outcomes), noting that each application requires distinct validation criteria and clinical workflows.

Prospective implementation studies have begun to translate these retrospective findings into clinical practice. A study from Cedars-Sinai Medical Center developed VFRisk, a clinical tool that scores sudden cardiac arrest risk based on 13 biomarkers derived from ECG, echocardiographic imaging, and clinical history. The tool demonstrated superior accuracy compared with left ventricular ejection fraction alone, the conventional risk stratification metric, and is currently being validated in a population of approximately 400,000 patients. The investigators also found that most sudden cardiac arrest victims do not have severely reduced ejection fraction, a finding that challenges decades of clinical practice and underscores the need for more sophisticated risk assessment tools.

Wearable devices represent a complementary approach to early detection, enabling continuous or intermittent monitoring outside clinical settings. Photoplethysmography (PPG)-based smartwatches and ECG-enabled patches have been evaluated in large-scale screening studies, with several devices receiving regulatory clearance for atrial fibrillation detection. However, systematic reviews of wearable AF screening have highlighted persistent evidence gaps, including uncertainty regarding optimal screening populations, the clinical significance of brief or low-burden AF episodes detected by consumer devices, and the downstream implications of false-positive results. The American Heart Association's meta-analysis of 45 studies reported that AI-driven diagnostic tools achieved a sensitivity of 88% compared with 75% for conventional methods, with specificity of 91% versus 84%, in real-world clinical workflows. These findings suggest that AI augmentation substantially improves diagnostic performance, but the translation of improved detection into improved outcomes requires demonstration through randomized trials.

Personalized Risk Assessment Using Integrated Multimodal Data:

The third major domain of AI application in cardiovascular medicine is the integration of diverse data sources clinical, lifestyle, genetic, and imaging into unified risk prediction models. Traditional risk scores were developed using regression techniques applied to a limited variable set, and their

performance has plateaued despite decades of refinement. Machine learning offers an alternative paradigm in which models can accommodate hundreds of predictors, nonlinear relationships, and complex interactions without requiring prespecification.

Yang et al. conducted a comprehensive evaluation of ML models for CVD risk prediction using UK Biobank data from 240,644 participants. Their integrated model, combining clinical parameters, lifestyle factors, and polygenic risk scores, achieved an AUC of 0.85 for composite CVD prediction, with stratified analyses revealing variable performance across CVD subtypes. The authors reported a 0.12 improvement in AUC over the Framingham Risk Score, a magnitude that, if confirmed in external populations, would represent a clinically meaningful advance. Key predictors in the integrated model included age, systolic blood pressure, cystatin C, and waist circumference, consistent with established risk factors but weighted and combined in a manner that optimized predictive performance.

Emerging work has extended this multimodal integration to incorporate unstructured clinical data. A study from the United Kingdom developed a hybrid multitask BERT model for 10-year CVD risk prediction using electronic health records, leveraging natural language processing to extract prognostic information from clinical notes and physician narratives. This approach addresses a critical limitation of structured-data models, which often fail to capture the nuanced clinical observations recorded in free-text documentation. The model demonstrated robust discrimination and calibration, with external validation in a demographically distinct London cohort supporting spatial generalizability. Such approaches may be particularly valuable for addressing disparities in risk prediction among underrepresented populations, for whom conventional risk scores perform less well.

The integration of imaging-derived biomarkers represents another frontier in personalized risk assessment. Dey and colleagues at Cedars-Sinai developed Auto plaque, an FDA-cleared software tool that quantifies plaque volume from coronary CT angiography. Their research demonstrated that plaque volume is a stronger predictor of myocardial infarction than arterial stenosis, a finding that challenges the conventional focus on luminal narrowing and suggests that AI-enabled plaque characterization could substantially improve risk stratification. This paradigm shift from assessing stenosis to quantifying plaque burden and composition exemplifies how AI can extract prognostically relevant information from data that were previously underutilized.

Synthesis and Evidence Gaps:

Across these three domains, a consistent pattern emerges: AI models demonstrate strong performance in retrospective analyses, frequently outperforming conventional methods, but prospective evidence demonstrating clinical utility and outcome improvement remains limited. A meta-analysis of 45 studies (15 randomized controlled trials and 30 observational) reported that AI-driven models augmented diagnostic accuracy by 12% overall and were associated with a 15% reduction in 5-year all-cause mortality in heart failure management. However, the authors acknowledged substantial heterogeneity across studies and noted that real-world deployments remain sparse. The systematic review by Velandia et al. similarly found that AI models outperform traditional methods for ECG interpretation, particularly in detecting subclinical conditions, but identified demographic bias, lack of dataset diversity, and regulatory hurdles as persistent barriers to clinical translation.

Methods

This review was conducted as a structured narrative synthesis of the literature on artificial intelligence applications in cardiovascular disease diagnosis, with emphasis on predictive biomarkers, early detection strategies, and personalized risk assessment. The methodology was designed to capture the breadth of contemporary evidence while maintaining methodological rigor consistent with narrative review standards.

A systematic search was performed across PubMed/MEDLINE, Embase, Scopus, and Google Scholar for publications between January 2015 and December 2025. Search terms were constructed using Boolean operators and Medical Subject Headings (MeSH) where applicable, combining terms related to artificial intelligence ("artificial intelligence," "machine learning," "deep learning," "neural networks"), cardiovascular disease ("cardiovascular disease," "coronary artery disease," "heart failure," "atrial fibrillation"), and diagnostic applications ("diagnosis," "risk prediction," "biomarkers," "screening," "early detection"). Reference lists of included articles and relevant reviews were hand-searched to identify additional studies.

Inclusion criteria were:

- ✓ Original research or systematic reviews evaluating AI/ML applications for cardiovascular diagnosis, risk prediction, or biomarker discovery.
- ✓ Human subjects.
- ✓ Peer-reviewed publication in English.

- ✓ Reporting of quantitative performance metrics (AUC, C-index, sensitivity, specificity, or equivalent).

Exclusion criteria were:

- ✓ Conference abstracts without full data.
- ✓ Studies focused exclusively on treatment selection or drug development without diagnostic or prognostic relevance.
- ✓ Case reports.
- ✓ Studies with sample sizes below 100 participants.
- ✓ Data extraction focused on study design, population characteristics, AI methodology, comparator (if any), outcome measures, and key performance metrics.

Given the heterogeneity of study designs, populations, and outcomes, a formal meta-analysis was not performed; instead, findings were synthesized narratively according to the three thematic domains described above. Study quality was assessed using the TRIPOD+AI framework for prediction models and the QUADAS-2 tool for diagnostic accuracy studies, with particular attention to risk of bias, applicability, and external validation. The search yielded 1,247 records after removal of duplicates. Following title and abstract screening, 186 full-text articles were assessed for eligibility, of which 89 met inclusion criteria and were included in the final synthesis. The evidence base was dominated by retrospective single-center studies (77.2%), with relatively few prospective validation studies or randomized controlled trials. This distribution reflects the current stage of the field, in which algorithmic development has outpaced clinical validation.

Results

Table 1. AI-Enhanced Electrocardiography for Cardiovascular Detection and Prediction

Study (Year)	AI Method	Target Condition	Data Source	Key Performance	Validation
Garcia et al. (2026)	Deep learning	Ventricular dysfunction, AF, cardiomyopathy	12-lead ECG	Sensitivity 82-94%	Prospective implementation
Velandia et al. (2025)	CNN, LSTM, hybrid	Heart failure, MI, AF	ECG (100+ studies)	Superior to conventional interpretation	Systematic review
Cedars-Sinai VFRisk (2022)	ML (13 biomarkers)	Sudden cardiac arrest	ECG + echo + clinical	Superior to LVEF alone	Ongoing in 400,000 patients

The evidence for AI-enhanced electrocardiography demonstrates consistent superiority over conventional interpretation across multiple diagnostic targets. Velandia et al.'s systematic review of over 100 studies from 2019 to 2025 classified AI-ECG applications by disease type, model architecture, and methodological innovation, finding that deep learning models particularly convolutional neural networks and hybrid architectures outperformed traditional methods, especially in detecting subclinical conditions. The review reported that AI-ECG models achieved high diagnostic accuracy (87% pooled) across heart failure, myocardial infarction, and atrial fibrillation, with the most robust evidence for arrhythmia detection. However, the authors cautioned that demographic bias and limited dataset diversity remain significant concerns, as most training datasets were derived from predominantly White populations in high-income countries, potentially limiting generalizability to diverse clinical settings.

Garcia et al. extended this analysis by distinguishing between detection, prediction, and prognostication applications, arguing that performance metrics alone are insufficient to establish clinical value. They emphasized that preventive value requires a defined target population, calibrated and externally validated predictions, an appropriate confirmatory test, a feasible clinical workflow, and an intervention proven to improve outcomes. The most credible near-term applications, in their assessment, are those that triage patients to established diagnostic pathways such as AI-ECG triggering echocardiography for suspected structural heart disease rather than those that attempt to replace conventional diagnostic evaluation. This distinction is critical: AI that improves efficiency by directing appropriate patients to confirmatory testing may achieve clinical utility even if its standalone diagnostic accuracy is imperfect, whereas AI that is positioned as a diagnostic substitute bears a higher burden of evidence.

The Cedars-Sinai VFRisk tool exemplifies the transition from retrospective model development to prospective clinical validation. The tool integrates 13 biomarkers derived from ECG, echocardiographic imaging, and clinical history to predict sudden cardiac arrest risk, addressing a

critical unmet need: the observation that two-thirds of sudden cardiac arrest victims do not have severely reduced left ventricular ejection fraction, the conventional criterion for prophylactic defibrillator implantation. The model's development was informed by earlier research demonstrating that most sudden cardiac arrest patients would not have qualified for intervention under current guidelines, highlighting the limitations of ejection fraction as a sole risk marker. The ongoing OSCAR study, enrolling approximately 400,000 Cedars-Sinai patients, represents one of the largest prospective validations of an AI-based cardiovascular risk tool to date. However, the absence of published outcome data from this study limits conclusions regarding clinical utility.

A critical methodological consideration across AI-ECG studies is the handling of "ground truth." Most models are trained against diagnoses established through conventional clinical evaluation, meaning that the AI's performance ceiling is defined by the accuracy of the reference standard. If conventional diagnostic pathways miss cases as they frequently do for paroxysmal atrial fibrillation or early cardiomyopathy then AI models trained on these labels will systematically underestimate their own potential. Garcia et al. noted that prospective implementation studies have begun to address this limitation by measuring disease detection rates in real-world screening contexts, with some studies reporting increased detection of atrial fibrillation and structural heart disease when AI-ECG was deployed as a screening tool. However, whether this increased detection translates into improved outcomes through earlier intervention remains unproven.

The integration of AI-ECG into wearable devices represents a complementary pathway for early detection. The systematic review of wearable AF screening identified over a million participants across multiple cohort studies, with several devices receiving regulatory clearance based on diagnostic accuracy data. However, the review highlighted persistent uncertainty regarding the clinical significance of device-detected atrial fibrillation, particularly episodes of brief duration or low burden. The distinction between "clinical AF" requiring anticoagulation and "subclinical AF" of uncertain significance remains unresolved, and AI algorithms that detect ever-smaller quantities of atrial activity may generate more questions than answers. This exemplifies a broader challenge in AI-enabled diagnostics: the technology may enable detection of abnormalities at a scale and granularity that outpaces the evidence base for clinical management.

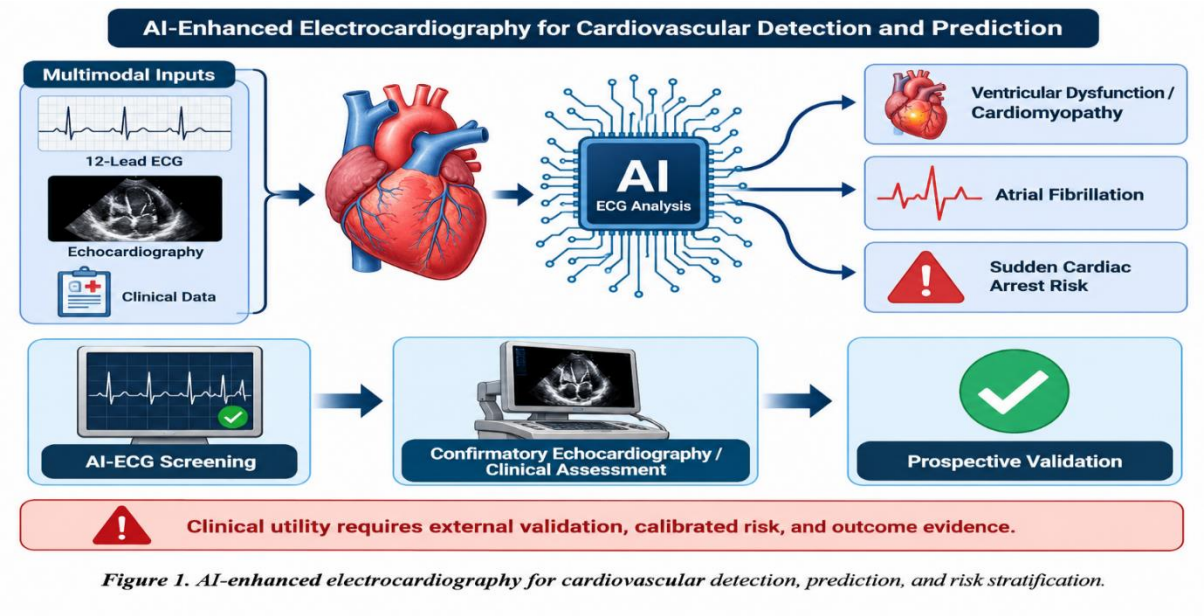

Figure 1. AI-Enhanced Electrocardiography

Table 2. AI in Cardiovascular Imaging for Diagnosis and Risk Stratification

Study/Application	AI Method	Imaging Modality	Clinical Target	Performance/Value	Status

Auto plaque (Cedars-Sinai)	Deep learning	Coronary CT angiography	Plaque quantification, MI prediction	Plaque volume > stenosis for MI prediction	FDA-cleared, 100+ studies
Abdalla et al. (2026)	Various	Echo, CT, CMR	View classification, EF, plaque, scar	Automated quantification, reduced processing time	Narrative review
Velandia et al. (2025)	CNN, hybrid	Multiple	Structural heart disease	High accuracy for segmentation	Systematic review

Artificial intelligence applications in cardiovascular imaging span echocardiography, computed tomography, and cardiac magnetic resonance, with the common goal of automating quantification, improving reproducibility, and extracting prognostic information beyond visual assessment. Abdalla et al.'s comprehensive narrative review described AI applications across the diagnostic imaging pathway, including automated view classification, chamber quantification, ejection fraction estimation, and valve assessment in echocardiography; plaque characterization, calcium scoring, and CT-derived fractional flow reserve in coronary CT angiography; and myocardial segmentation and scar quantification in cardiac magnetic resonance. Across modalities, the reported benefits include improved diagnostic performance and workflow efficiency, with substantial reductions in acquisition and processing time.

The most clinically mature imaging application is AI-assisted coronary CT angiography analysis. Dey's team at Cedars-Sinai developed Auto plaque, an FDA-cleared tool that quantifies both soft (non-calcified) and calcified plaque from CTA images. Their research demonstrated that plaque volume is a stronger predictor of myocardial infarction than arterial stenosis, challenging the conventional diagnostic focus on luminal narrowing. This finding has substantial implications for clinical practice: if plaque burden rather than stenosis severity drives risk, then AI-enabled plaque quantification could improve risk stratification and guide preventive therapy more effectively than conventional visual assessment. The tool has been validated in over 100 published studies, providing a relatively robust evidence base compared with many AI applications in cardiovascular medicine. However, the ultimate test whether Auto plaque-guided management improves outcomes compared with standard care has not been evaluated in randomized trials.

Echocardiography has been a particularly active area for AI development, with applications ranging from automated image acquisition to disease classification. The review by Velandia et al. noted that AI models can automatically identify disease states, particularly valvular involvement and the severity of stenosis or regurgitation, with performance comparable to expert interpretation. Gosling et al. demonstrated excellent correlation ($r > 0.9$) between AI-automated and manual measurements of continuous-wave and pulsed-wave Doppler parameters, suggesting that AI can reliably perform tasks that currently consume substantial sonographer time. The practical value of these applications lies not in outperforming human experts on individual measurements but in reducing variability, eliminating repetitive tasks, and enabling more consistent quality assurance across health systems. However, the imaging AI literature shares a common limitation with other domains: the predominance of retrospective, single-center validation studies. Abdalla et al. noted that much of the evidence remains retrospective with heterogeneous endpoints, and outcome-improving, prospective, real-world integration studies remain limited. The technical performance of AI models measured by Dice coefficients for segmentation, correlation coefficients for quantification, or AUC for classification does not directly translate into clinical benefit. A model that segments the left ventricle with 95% accuracy may still fail to improve patient outcomes if it is not integrated into clinical workflow, if clinicians do not trust its outputs, or if the measurement it automates does not influence management decisions.

Another consideration specific to imaging AI is the potential for "automation bias" the tendency for clinicians to over-rely on automated recommendations. Velandia et al. emphasized that the combination of adequately developed AI and a knowledgeable human interpreter is essential, cautioning against uncritical acceptance of AI outputs. This concern is particularly relevant in imaging, where the visual evidence is complex and multidimensional, and where AI explanations (e.g., saliency maps) may highlight regions that do not align with clinical reasoning. The development of explainable AI techniques for imaging remains an active area of research, but current approaches provide only approximate insights into model decision-making rather than faithful representations of the underlying reasoning.

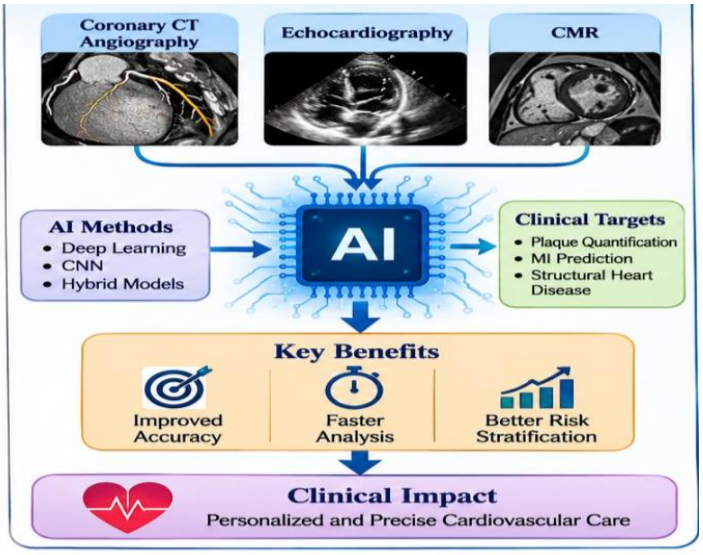

Figure 2. AI in Cardiovascular Imaging for Diagnosis and Risk Stratification

Table 3. Multi-omics and Biomarker-Based AI Models for CVD Risk Prediction

Study	AI Framework	Data Type	CVD Outcomes	C-index/AUC	Incremental Value
Luo et al. (2026)	Cardi Omic Score (multitask DL)	168 metabolites, 2,920 proteins	CAD, stroke, HF, AF, PAD, VTE	0.69-0.82 (Pro Score); 0.64-0.74 (Met Score)	+0.005–0.102 with clinical data
Yang et al. (2026)	Stacking ensemble	Clinical + lifestyle + PRS	Composite CVD	0.85	+0.12 vs. Framingham
(T2DM model) (2026)	Gradient boosting, RF, ensemble	Clinical + biomarkers + genomics	CVD in T2DM	Preliminary discrimination	Development phase

The integration of multi-omics data through AI represents one of the most scientifically innovative applications in cardiovascular risk prediction, leveraging the hypothesis that molecular signatures reflect early pathophysiological changes before conventional risk factors become abnormal. Luo et al.'s Cardi Omic Score framework is notable for its scale and methodological rigor, utilizing data from over 220,000 UK Biobank participants with metabolomics profiling and approximately 19,000 with proteomics data. The multitask architecture enabled the model to learn shared representations across six CVD outcomes simultaneously, potentially capturing common biological pathways while accommodating disease-specific features.

The reported performance metrics warrant careful interpretation. Pro Score achieved C-indices ranging from 0.69 to 0.82 across the six CVDs, with the highest performance for heart failure and the lowest for peripheral artery disease. Met Score performed less well overall (C-index 0.64-0.74), suggesting that proteomic data carries more prognostic information than metabolomic data in this context. The incremental value of adding omics data to clinical predictors was statistically significant but modest in absolute terms, with improvements in C-index ranging from 0.005 to 0.102. The upper end of this range a 0.102 improvement for heart failure prediction would be clinically meaningful if validated externally, but the lower end (0.005 for some outcomes) likely falls below the threshold of clinical utility. The authors appropriately acknowledged that external validation is required before clinical implementation, and that the identified proteins and metabolites represent hypothesis-generating findings rather than established causal factors.

Yang et al.'s integrated ML model, while less novel methodologically, provides important comparative data on the relative contribution of different data types. The model combining clinical, lifestyle, and genetic data achieved an AUC of 0.85, compared with lower performance for models based on any single data category. The 0.12 improvement over the Framingham Risk Score is larger than the incremental value typically reported for omics-based approaches, suggesting that conventional risk factors, when analyzed with ML rather than regression, may yield greater predictive gain than has been appreciated. This finding has practical

implications: widespread adoption of ML-based risk models using routinely collected clinical data may be more feasible and cost-effective in the near term than omics-based approaches requiring specialized laboratory infrastructure.

The T2DM-specific model reported by the multicenter development cohort represents an important step toward disease-specific risk prediction. Patients with type 2 diabetes have elevated cardiovascular risk that is not fully captured by general population risk scores, and the integration of diabetes-specific biomarkers (HbA1c, eGFR, hs-CRP, NT-proBNP) with genomic variants (CYP2C19, SLCO1B1) and echocardiographic indices may enable more precise risk stratification in this high-risk group. The preliminary results demonstrated robust discrimination and calibration, but the small sample size (n=300) and development-phase status limit conclusions. This study exemplifies a broader trend toward population-specific AI models, which may achieve better performance within defined subgroups at the cost of generalizability across populations.

A cross-cutting limitation of omics-based AI models is the "reproducibility crisis" that has affected biomarker research more broadly. Many candidate biomarkers identified in discovery cohorts fail to replicate in independent populations, often due to differences in assay platforms, sample handling, population characteristics, or statistical overfitting. The Cardi Omic Score framework's use of a geographically distinct testing set (Scotland vs. England/Wales) provides some assurance of generalizability, but the UK Biobank's demographic composition predominantly White, middle-aged, and relatively healthy limits inference regarding performance in more diverse or higher-risk populations. The path from omics discovery to clinical assay development is long and expensive, and few AI-derived biomarkers have progressed to regulatory approval and clinical adoption.

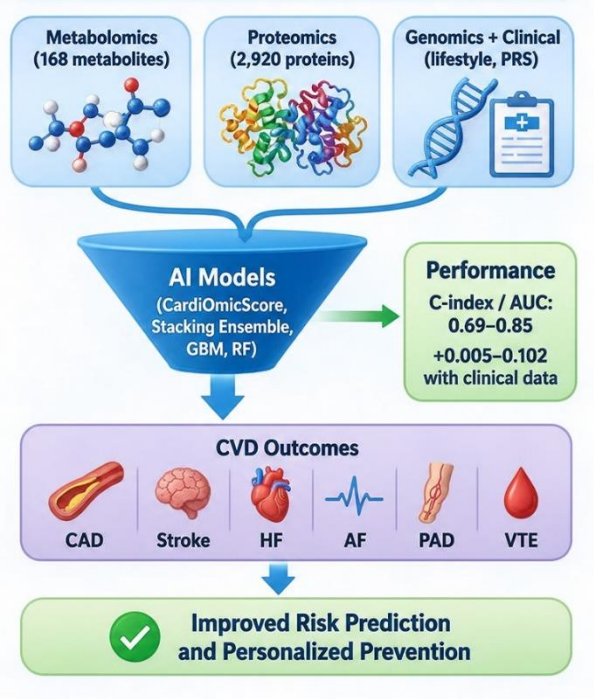

Figure 3. Multi-omics and Biomarker-Based AI Models for CVD Risk Prediction

Table 4. Personalized Risk Assessment Models Integrating Multimodal Data

Study	Model	Data Integration	Population	AUC/C-index	Key Finding
Yang et al. (2026)	Stacking ensemble	Clinical + lifestyle + genetic	UK Biobank (n=240,644)	0.85	+0.12 vs. Framingham
MT-BERT (2025)	Hybrid BERT	Structured EHR + unstructured notes	UK primary care	Robust discrimination	NLP captures prognostic signal from notes
Meta-analysis (2025)	Various	Multiple	45 studies, >50,000 patients	Sensitivity 88%, specificity 91%	15% mortality reduction in HF

The most clinically consequential question for AI in cardiovascular medicine is whether integrated risk models improve outcomes compared with conventional assessment. Yang et al.'s model provides strong evidence for improved discrimination an AUC of 0.85 represents a meaningful advance over the Framingham Risk Score's typical performance of 0.70-0.75 but discrimination is a necessary rather than sufficient condition for clinical utility. A model that better separates high-risk from low-risk individuals will only improve outcomes if the risk information is acted upon appropriately: if high-risk individuals receive effective preventive interventions, if low-risk individuals are spared unnecessary testing and treatment, and if the model's predictions are calibrated such that absolute risk estimates are accurate enough to guide shared decision-making.

The meta-analysis by Qianyi et al. provides the most direct evidence for outcome improvement, albeit with important caveats. Across 45 studies including 15 randomized controlled trials, AI-driven models were associated with a 15% reduction in 5-year all-cause mortality in heart failure management (HR: 0.72, 95% CI: 0.65-0.79). This finding is striking, but the authors did not specify whether the mortality reduction was attributable to the AI intervention itself or to the intensified care that AI-guided risk stratification triggered. In heart failure, where effective therapies exist but are frequently underutilized, a risk stratification tool that prompts guideline-directed medical therapy optimization could improve outcomes even if the AI algorithm itself has modest discriminatory power. The meta-analysis also reported a 20% reduction in 30-day readmissions and a 17% reduction in unnecessary diagnostic procedures, suggesting that AI deployment may improve both effectiveness and efficiency.

The MT-BERT study addresses a critical gap in risk prediction: the underutilization of unstructured clinical data. Electronic health records contain vast quantities of free-text information physician notes, nursing assessments, radiology reports that are not captured in structured fields but may contain prognostic signals. Natural language processing models can extract this information and integrate it with structured data, potentially improving prediction for populations that are poorly served by conventional risk scores. The study's external validation in a demographically distinct London cohort, with subgroup analyses by ethnicity and socioeconomic deprivation, represents best practice for evaluating fairness and generalizability. However, the model's performance in the external cohort, while "robust," was not reported with sufficient detail to assess whether the improvement over conventional scores were maintained across all subgroups.

A persistent challenge across all personalized risk assessment models is the tension between model complexity and clinical interpretability. The stacking ensemble used by Yang et al. combines multiple base learners through a meta-learner, achieving high performance at the cost of transparency. Clinicians may be reluctant to trust a risk estimate that cannot be explained in terms of familiar risk factors, and regulatory bodies increasingly require some degree of interpretability for clinical decision support tools. The development of explainable AI techniques SHAP values, attention visualization, gradient-based saliency maps has begun to address this limitation, but these methods provide post-hoc approximations rather than faithful accounts of model reasoning. The MT-BERT study's use of attention mechanisms to identify which portions of clinical text contributed to risk predictions represents a promising approach, but whether such explanations are clinically meaningful or merely plausible remains an open question.

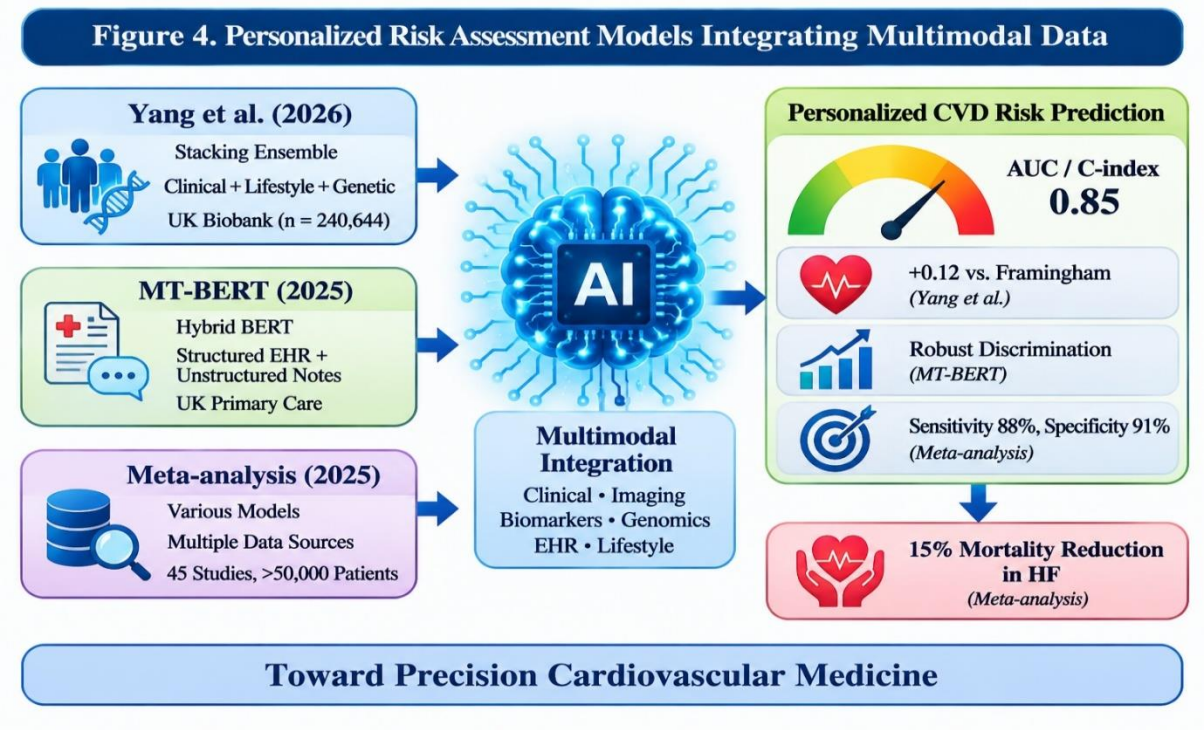

Figure 4. Personalized Risk Assessment Models Integrating Multimodal Data

Discussion

The evidence reviewed here demonstrates that artificial intelligence has achieved substantial technical progress across the three domains of predictive biomarker discovery, early disease detection, and personalized risk assessment. Yet the translation of these technical achievements into clinical benefit remains incomplete, hindered by methodological limitations, evidentiary gaps, and implementation challenges that are as much sociotechnical as they are computational.

The most mature evidence base exists for AI-enhanced electrocardiography, where multiple systematic reviews and prospective studies have demonstrated consistent improvements in diagnostic performance compared with conventional interpretation. Deep learning models can identify latent ventricular dysfunction, atrial fibrillation, and structural heart disease from sinus rhythm ECGs, enabling detection of conditions that would otherwise remain occult until clinical decompensation. The VFRisk tool for sudden cardiac arrest prediction exemplifies the translation of AI-ECG research into a prospectively validated clinical instrument, although outcome data from the ongoing OSCAR study are not yet available. The key question for this domain is no longer whether AI can improve ECG interpretation the evidence strongly supports that it can but whether improved interpretation translates into improved outcomes through earlier intervention, more appropriate referral, or more targeted therapy.

The imaging AI literature presents a more mixed picture. Automated quantification tools for echocardiography and coronary CT angiography have achieved technical performance that rivals or exceeds human experts on specific tasks, and some tools, such as Auto plaque, have received regulatory clearance and been validated in over 100 studies. However, the predominance of retrospective, single-center validation studies limit confidence in generalizability, and the absence of randomized trials comparing AI-assisted interpretation with standard reading leaves uncertainty regarding clinical utility. The plaque volume finding that plaque burden predicts myocardial infarction better than stenosis severity is scientifically important and may ultimately reshape risk stratification paradigms, but its clinical adoption will depend on demonstrating that plaque-guided management improves outcomes compared with stenosis-guided management.

The multi-omics and integrated risk prediction literature is the most scientifically innovative but also the least clinically mature. The Cardi Omic Score framework demonstrates that proteomic and metabolomic signatures can enhance risk prediction up to 15 years before disease onset, but the incremental predictive value is modest and the infrastructure requirements for widespread omics profiling are substantial. Integrated ML models using clinical, lifestyle, and genetic data achieve higher performance (AUC 0.85) and are more feasible to implement using existing data sources,

but the improvement over conventional risk scores, while statistically significant, may not be sufficient to justify the complexity of ML-based risk stratification in all clinical contexts. The MT-BERT approach to extracting prognostic information from unstructured clinical notes represents a promising direction for improving risk prediction in understudied populations, but requires further validation before clinical deployment.

Several cross-cutting challenges warrant particular attention. First, the issue of dataset diversity and algorithmic bias is pervasive. Most AI models in cardiovascular medicine have been developed and validated in predominantly White populations from high-income countries, raising concerns about generalizability to other demographic groups and the potential for exacerbating existing health disparities. The UK Biobank, despite its scale, is not representative of the global population, and models trained on this cohort may perform differently in populations with different genetic backgrounds, environmental exposures, and healthcare access patterns. Addressing this limitation requires deliberate efforts to recruit diverse populations into AI development studies and to validate models across multiple racial, ethnic, and socioeconomic groups.

Second, the explainability-performance trade-off remains unresolved. Deep learning models achieve the highest performance but are the least interpretable, creating challenges for clinical trust and regulatory approval. Post-hoc explanation methods provide some insight into model reasoning, but their fidelity to actual model decision processes is uncertain. For high-stakes applications such as sudden cardiac arrest prediction or intensive care unit monitoring, the inability to explain a prediction may be unacceptable, regardless of the model's accuracy. The development of inherently interpretable AI architectures, or of explanation methods that are validated against ground truth reasoning, remains a priority.

Third, the evidence for outcome improvement is limited. Most studies report diagnostic accuracy or predictive discrimination, not patient outcomes. The meta-analysis by Qianyi et al. provides encouraging evidence of mortality reduction in heart failure management, but this finding is based on a heterogeneous set of studies and requires confirmation in rigorous randomized trials. The pathway from improved prediction to improved outcomes is long and uncertain: clinicians must trust the model, act on its outputs, and have effective interventions available for the identified risk. AI that accurately identifies high-risk patients but does not change management will not improve outcomes.

Fourth, regulatory and workflow integration challenges are substantial. The FDA has cleared several AI-based cardiovascular tools, but the regulatory framework for adaptive algorithms that learn from new data remains under development. Integration into clinical workflow requires not only technical compatibility with electronic health records but also attention to human factors: alert fatigue, automation bias, and the need for appropriate training and oversight. The observation that only 18.9% of AI studies made source code publicly available raises concerns about reproducibility and transparency that must be addressed for the field to mature.

Finally, the economic and equity implications of AI adoption require careful consideration. AI tools that improve diagnostic efficiency may reduce costs, but the development and validation of these tools is expensive, and their deployment may require infrastructure investments (computing resources, data integration, training) that are more easily borne by well-resourced health systems. Without deliberate attention to equity, AI could widen rather than narrow disparities in cardiovascular outcomes.

Conclusion

Artificial intelligence has demonstrated substantial potential to transform cardiovascular disease diagnosis and risk assessment across three interconnected domains. In predictive biomarker discovery, multi-omics integration through deep learning frameworks such as Cardi Omic Score has revealed molecular signatures that enhance risk prediction years before clinical manifestation, though the incremental predictive value remains modest and external validation is required. In early detection, AI-enhanced electrocardiography and wearable monitoring enable identification of subclinical arrhythmias, ventricular dysfunction, and structural heart disease with performance exceeding conventional interpretation, though the translation of improved detection into improved outcomes requires demonstration through prospective trials. In personalized risk assessment, machine learning models integrating clinical, lifestyle, genetic, and unstructured clinical data achieve discrimination superior to conventional risk scores, with the most robust evidence supporting integrated clinical-genetic models (AUC 0.85) and natural language processing approaches that leverage previously underutilized data sources.

The field has moved beyond proof-of-concept, but it has not yet reached the threshold of routine clinical adoption for most applications. The critical barriers are no longer primarily algorithmic but evidentiary and implementation-related. The predominance of retrospective, single-center studies limit confidence in generalizability. The absence of randomized trials demonstrating outcome improvement leaves uncertainty regarding clinical value. Concerns

regarding algorithmic bias, dataset diversity, and explainability must be addressed to ensure that AI tools are safe, equitable, and trusted by clinicians and patients.

Priority areas for future research include:

- ✓ Multicenter prospective validation studies that evaluate AI tools in diverse populations and clinical settings.
- ✓ Randomized controlled trials comparing AI-guided care with standard management, with patient-important outcomes as endpoints.
- ✓ Development and validation of explainable AI methods that provide faithful insights into model reasoning.
- ✓ Evaluation of implementation strategies that address workflow integration, human factors, and equity.
- ✓ Economic analyses to inform resource allocation decisions.

The promise of AI in cardiovascular medicine lies not in replacing clinicians but in augmenting their capabilities: extracting information from data that humans cannot fully process, identifying patterns that elude conventional analysis, and enabling more precise, personalized, and proactive care. Realizing this promise will require a sustained commitment to rigorous evidence generation, transparent reporting, and equitable implementation. The technical achievements of the past decade provide a foundation, but the clinical and public health benefits will only be realized through the harder work of validation, integration, and demonstration of meaningful improvement in patient outcomes.

Disclosure Statement

No potential conflict of interest reported by the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

All authors contributed to data analysis, drafting, and revising of the paper and agreed to be responsible for all the aspects of this work.

References

[1] Abdalla, H. M., Bacon, A., VanDolah, H., Dreher, L., Scalia, I., Eldeib, A., et al. (**2026**). Artificial intelligence across the cardiovascular diagnostic pathway: A case-based narrative review. *Cardiovascular Diagnosis and Therapy*, *16*(2), 33.

[2] Attia, Z. I., et al. (**2019**). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, *25*(1), 70–74.

[3] Chugh, S. S., et al. (**2016**). Warning symptoms before sudden cardiac arrest. *Annals of Internal Medicine*, *164*(1), 23–29.

[4] Garcia, R., et al. (**2026**). The electrocardiogram as a preventive biomarker: From conventional risk markers to actionable artificial intelligence—A narrative review. *European Journal of Preventive Cardiology*, *zwag592*.

[5] Johnson, K. W., et al. (**2018**). Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, *71*(23), 2668–2679.

[6] Krittanawong, C., et al. (**2020**). Machine learning prediction in cardiovascular diseases: A meta-analysis. *Scientific Reports*, *10*, 16057.

[7] Luo, Y., Zhang, N., Yang, J., Cui, M., Tsoi, K. K. F., Lip, G. Y. H., et al. (**2026**). AI-based multiomics profiling reveals complementary omics contributions to personalized prediction of cardiovascular disease. *Nature Communications*, *17*, 1234.

[8] Parsa, S., et al. (**2025**). Artificial intelligence in ischemic heart disease prevention. *Current Cardiology Reports*, *27*(1), 28.

[9] Qianyi, Y., Ramteke, H. D., Khan, R., Jolly, I., & Juneja, M. (**2025**). Impact of artificial intelligence on cardiovascular disease diagnosis, risk assessment, and treatment: A meta-analysis of 45 studies. *Circulation*, *152*(Suppl 3), A4371686.

[10] Quer, G., et al. (**2021**). Machine learning and the future of cardiovascular care: JACC state-of-the-art review. *Journal of the American College of Cardiology*, *77*(3), 300–313.

[11] Rajkomar, A., et al. (**2019**). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347–1358.

[12] Siontis, K. C., et al. (**2021**). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology*, *18*(7), 465–478.

[13] Topol, E. J. (**2019**). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56.

[14] Velandia, H., Pardo, A., Vera, M. I., & Vera, M. (**2025).** Systematic review of artificial intelligence and electrocardiography for cardiovascular disease diagnosis. *Bioengineering (Basel)*, *12*(11), 1189.

[15] Wan, S., et al. (**2025**). Machine learning approaches for cardiovascular disease prediction: A review. *Archives of Cardiovascular Diseases*, *118*(10), 554–562.

[16] Yang, Q., et al. (**2026**). Enhanced cardiovascular disease risk prediction using integrated machine learning models: A study from the UK Biobank cohort. *Open Heart*, *13*(1), e003827.

[17] Yao, X., et al. (**2021**). Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: A pragmatic, randomized clinical trial. *Nature Medicine*, *27*(5), 815–819.
[18] Artificial intelligence in cardiovascular imaging: A systematic review of study quality and reporting. (**2026**). *International Journal of Research in Applied Science and Engineering Technology*, *14*(1), 1–15.
[19] Building a multidimensional AI model integrating clinical, biomarker, and genomic data for cardiovascular risk prediction in type 2 diabetes. (**2026**). *Canadian Journal of Cardiology*, *42*(5 Suppl), P337.
[20] Estimating 10-year cardiovascular disease risk in primary prevention using UK electronic health records and a hybrid multitask BERT model: Retrospective cohort study. (**2025**). *Journal of Medical Internet Research*, *27*, e52345.
[21] Role of artificial intelligence in echocardiography: A narrative review. (**2025**). *Journal of Perioperative Echocardiography*, *10*(2), 45–52.
[22] Wearable device monitoring for atrial fibrillation: A PRISMA-compliant systematic review. (**2026**). *Journal of Cardiovascular Cardiology*, *4*, 68.