International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

REVIEW

AI-ECG for Echocardiography Triage in Structural Heart Disease: Evidence, Implementation, and Future Directions

Qianwen Tang[1,*], Kunfei Deng[2,*], Yu Cui[3], Hongjie Liu[2], Bairui Qian[1]

[1]Department of Cardiac Surgery, the First Hospital of China Medical University, Shenyang, 110001, People's Republic of China; [2]Department of Pancreatic-Biliary Surgery, the First Hospital of China Medical University, Shenyang, 110001, People's Republic of China; [3]Department of Cardiology, the First Hospital of China Medical University, Shenyang, Liaoning, 110001, People's Republic of China

*These authors contributed equally to this work

Correspondence: Hongjie Liu, Department of Pancreatic-Biliary Surgery, the First Hospital of China Medical University, No. 155 Nanjing North Street, Heping District, Shenyang, Liaoning, People's Republic of China, Tel +86-13889372639, Email hongjieliu_sy@yeah.net; Bairui Qian, Department of Cardiac Surgery, the First Hospital of China Medical University, No. 155 Nanjing North Street, Heping District, Shenyang, Liaoning, People's Republic of China, Tel +86-15541560177, Email qianbr1990@126.com

Abstract: Structural heart disease (SHD), including left ventricular systolic dysfunction, valvular heart disease, hypertrophic cardiomyopathy, cardiac amyloidosis, and pulmonary hypertension, remains underdiagnosed despite the increasing availability of disease-modifying therapies. Echocardiography is the principal confirmatory test, but its broad use as a screening tool is constrained by imaging capacity, cost, and referral efficiency. This review evaluates artificial intelligence-enabled 12-lead electrocardiography (AI-ECG) as a pre-echocardiographic triage tool for SHD. We synthesize evidence across reduced left ventricular ejection fraction, valvular disease, hypertrophic cardiomyopathy, cardiac amyloidosis, pulmonary hypertension, and composite SHD models, and distinguish two intended-use orientations: safety-net screening, in which a positive AI-ECG result serves as an additive trigger for confirmatory evaluation, and gatekeeper triage, in which a negative or low-risk AI-ECG result may support deferring or de-prioritizing echocardiography in selected low-risk settings. Current evidence most strongly supports low-LVEF detection, where pragmatic randomized implementation and early economic data are available. Valvular and composite SHD models are promising for referral enrichment, whereas hypertrophic cardiomyopathy, cardiac amyloidosis, and pulmonary hypertension remain earlier or pathway-incomplete applications. We also review false-positive interpretation, stepwise confirmation with point-of-care ultrasound, threshold selection, workflow integration, equity, regulation, and health economics. Overall, AI-ECG is currently best positioned as an additive safety-net tool to improve case finding upstream of echocardiography. Gatekeeper use remains investigational and requires prospective pathway-level validation, calibration, and operational safeguards before routine imaging deferral can be justified.

Keywords: AI-enabled electrocardiography, structural heart disease, echocardiography, safety-net screening, gatekeeper triage, surveillance

Introduction

The Undiagnosed Burden of Structural Heart Disease

Structural heart disease (SHD) encompasses a heterogeneous set of disorders, including left ventricular systolic dysfunction, valvular heart disease, hypertrophic cardiomyopathy (HCM), cardiac amyloidosis, and pulmonary hypertension (PH). Several of these phenotypes are either common or underrecognized and carry substantial morbidity and mortality.[1,2] Early detection of structural heart disease is clinically important, but broad imaging-based screening remains constrained by the cost and availability of echocardiography.[3] Asymptomatic left ventricular systolic dysfunction is estimated to affect approximately 3–6% of the general population and is associated with a substantially increased risk of progression to overt heart failure.[4] Aortic stenosis, the most prevalent valvular heart disease in developed countries, often

https://doi.org/10.2147/IJGM.S610298
Received: 18 March 2026
Accepted: 12 May 2026
Published: 29 May 2026

remains asymptomatic for years, but prognosis worsens sharply once symptoms develop.[5] Cardiac amyloidosis, once considered rare, is now recognized as an underdiagnosed and clinically important cause of heart failure with preserved ejection fraction, and substantial diagnostic delay remains common.[2,6]

Because effective disease-modifying therapies are now available for selected structural heart disease phenotypes—including obstructive HCM and transthyretin amyloid cardiomyopathy—the clinical value of timely diagnosis has increased.[7,8] Echocardiography remains the principal confirmatory imaging modality for most forms of structural heart disease, but broad referral strategies are difficult to scale because imaging capacity, expertise, and patient selection all constrain efficiency.[3,9] Accordingly, many patients remain undiagnosed until disease is advanced, and electrocardiogram (ECG)-based prediction models may help enrich referral toward higher-risk patients in a more targeted manner.[9]

Reconceptualizing the ECG as a Structural Screening Tool

The 12-lead ECG is among the most widely used cardiovascular tests worldwide.[10] Traditionally, its clinical role has centred on rhythm, conduction, ischemia, and indirect markers of chamber remodelling rather than deliberate screening for structural heart disease. Deep-learning studies have since shown that ECG waveforms contain latent information relevant to structural phenotyping beyond human visual interpretation.[11] A landmark 2019 study demonstrated that artificial intelligence (AI)-ECG could identify reduced left ventricular ejection fraction (LVEF) from routine 12-lead recordings with strong discriminatory performance, establishing the principle that an inexpensive and widely available test could function as an upstream screening instrument for abnormalities ordinarily confirmed by echocardiography.[12]

The conceptual shift was crystallized by EchoNext, a deep learning model trained on 1.2 million ECG–echocardiogram pairs from 230,018 unique patients across a diverse eight-hospital system. EchoNext explicitly framed AI-ECG deployment along a spectrum ranging from a "safety-net" strategy that triggers additional echocardiography to a later-stage "gatekeeper" strategy intended to reduce unnecessary imaging.[3] In operational terms, safety-net use refers to additive case-finding, in which a positive AI-ECG result prompts clinical review and confirmatory imaging for patients who might otherwise be missed. Gatekeeper use refers to rule-out or de-prioritization, in which a negative or low-risk AI-ECG result may support deferring echocardiography only in selected patients without independent imaging indications. This framing shifts the emphasis from AI-ECG as a diagnostic classifier in isolation to AI-ECG as a pre-echocardiographic triage tool embedded within a clinical pathway.

Aim, Novelty, and Real-World Relevance

The aim of this review is to synthesize current evidence on AI-ECG as a pre-echocardiographic triage tool for structural heart disease and to evaluate where the field is ready for safety-net deployment versus where it remains insufficient for gatekeeper use. A broad review has summarized the evolution of AI-enhanced electrocardiography in cardiovascular disease management, and a recent editorial has highlighted the promise of scalable AI-ECG screening for structural heart disease.[11,13] Building on this literature, the present review focuses on AI-ECG as a pre-echocardiographic triage intervention and shifts the discussion from model discrimination alone to pathway-level clinical use: when a positive AI-ECG should trigger confirmatory evaluation, when a negative or low-risk AI-ECG may support imaging deferral, and which validation standards are required for each intended-use scenario.

The review also synthesizes the core translational challenges of the field—evidence standards, calibration and thresholding, workflow accountability, equity and transportability, and lifecycle governance—through the same safety-net versus gatekeeper lens. We therefore compare reduced LVEF, valvular heart disease, hypertrophic cardiomyopathy, cardiac amyloidosis, pulmonary hypertension, and composite SHD models according to evidence maturity, referral yield, prospective validation, threshold implications, and deployment readiness. The real-world relevance is that AI-ECG is most likely to affect care not by replacing echocardiography, but by improving patient selection, prioritizing limited imaging capacity, and identifying patients who might otherwise remain undiagnosed. This review therefore frames AI-ECG as a clinical workflow intervention upstream of echocardiography and clarifies the asymmetric evidentiary requirements for safety-net screening versus gatekeeper triage.

Evidence Identification and Evidence Mapping

To improve methodological transparency and clarify the scope of the evidence base, we performed a structured evidence-mapping search focused on peer-reviewed studies evaluating AI-enabled electrocardiography for echocardiography-detectable structural heart disease or pre-echocardiographic triage. Records were identified through database searching in PubMed, Embase, and Web of Science, with supplementary sources including Google Scholar and citation tracking from reference lists. After deduplication and staged screening of titles, abstracts, and full texts, 26 core evidence publications were retained for qualitative evidence synthesis (Figure 1).

The included core evidence publications were selected because they directly evaluated AI-ECG model performance, external validation, prospective or randomized implementation, referral enrichment, health-economic evaluation, longitudinal surveillance, or pathway-level implications for structural heart disease triage. These publications covered reduced LVEF, valvular heart disease, hypertrophic cardiomyopathy, cardiac amyloidosis, pulmonary hypertension, composite SHD models, and stepwise confirmation workflows. Background epidemiology papers, treatment-context references, narrative reviews, methodological papers, fairness/regulatory literature, and adjacent non-AI-ECG studies were cited where relevant but were not counted as core evidence publications. Table 1 summarizes the key characteristics of the included core evidence publications, and Table 2 compares evidence maturity and deployment implications across SHD subtypes and pathways.

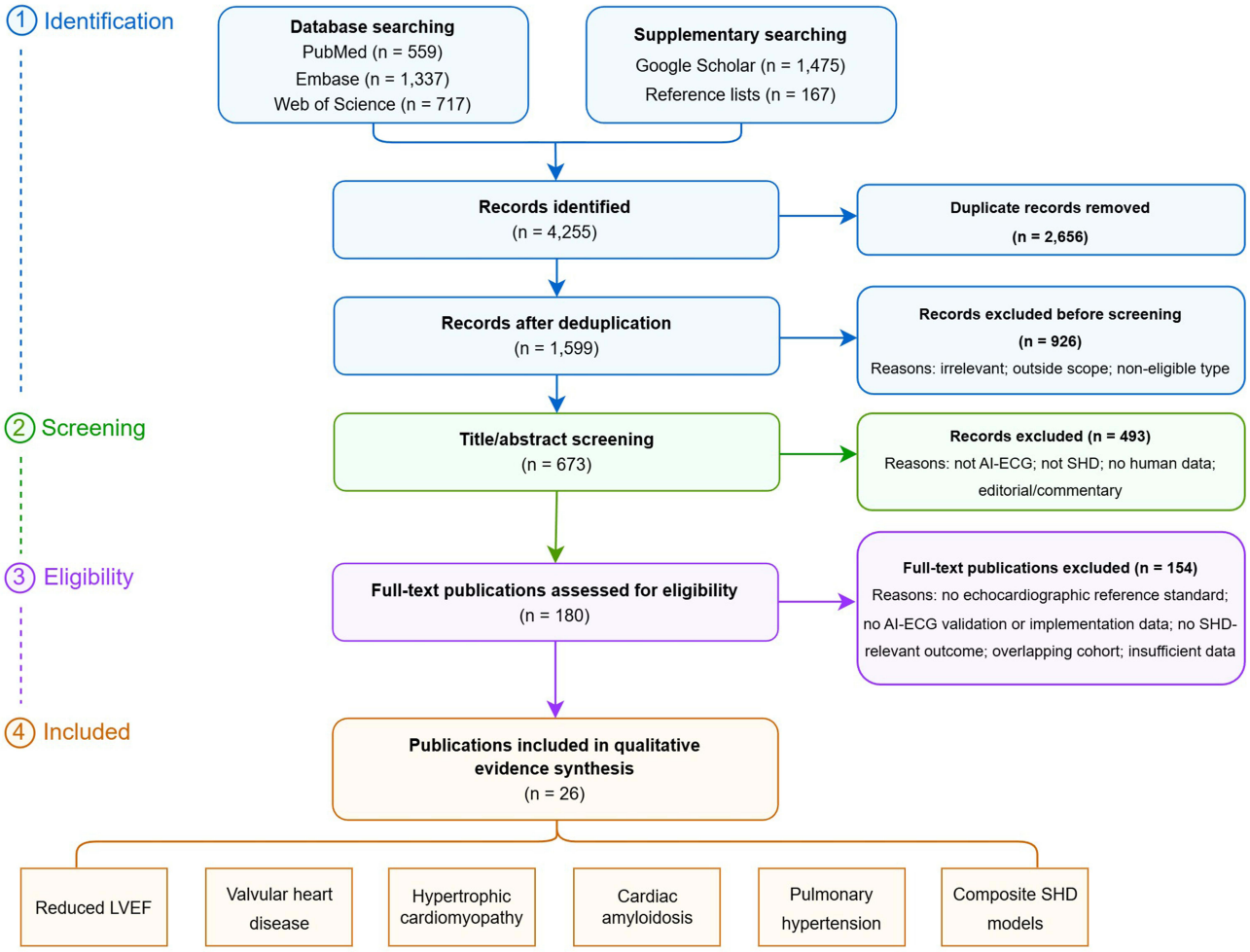

Figure 1 Evidence identification and selection process for AI-ECG studies in structural heart disease. This PRISMA-style flow diagram summarizes the structured identification and selection process for core AI-ECG evidence publications included in the qualitative evidence synthesis. The final evidence set comprised 26 core publications evaluating AI-ECG for reduced LVEF, valvular heart disease, hypertrophic cardiomyopathy, cardiac amyloidosis, pulmonary hypertension, composite SHD models, or related triage and surveillance pathways. The disease-domain boxes indicate the main evidence domains represented in the included evidence set and are not intended as mutually exclusive publication counts. Contextual references cited for background, implementation, ethics, regulation, or adjacent methodology were not counted as core evidence publications.

https://doi.org/10.2147/IJGM.S610298

https://doi.org/10.2147/IJGM.S610298

Table 1 Characteristics of Core Evidence Publications Included in the Qualitative Evidence Synthesis

Study	Target Phenotype/ Pathway	AI Input/Model or Strategy	Development or Evaluation Data	Validation Type	Key Reported Performance or Pathway Result	Evidence Phase
Attia et al, 2019[12]	Reduced LVEF	12-lead ECG; convolutional neural network	Large routine ECG-echocardiography cohort	Retrospective derivation and validation	AUC 0.93; NPV 98.7%; AI-positive/echo-negative patients had higher future ventricular dysfunction risk	Foundational model development
Yao et al, 2021; EAGLE[13]	Low EF	AI-ECG alert embedded in routine primary care	22,641 adults; 120 primary-care teams from 45 clinics/hospitals	Pragmatic cluster-randomized trial	New low-EF diagnosis increased from 1.6% to 2.1%; OR 1.32; P = 0.007	Randomized implementation
Tsai et al, 2025[14]	Low EF	AI-assisted ECG diagnosis/ prognostication	Inpatient clinical setting	Pragmatic randomized trial	Improved low-EF detection without increasing overall echocardiography utilization	Setting-specific implementation
Adedinsewo et al, 2024[15]	Peripartum cardiomyopathy/ cardiomyopathy screening	AI-assisted screening including digital-stethoscope and ECG pathways	Obstetric population in Nigeria	Pragmatic randomized trial	Diagnostic gain was mainly driven by the digital-stethoscope pathway; the stand-alone 12-lead AI-ECG signal was directionally concordant but less definitive	Low-resource prospective implementation
Rushlow et al, 2022[16]	Low EF	AI-ECG clinical decision support	Primary-care implementation cohort	Implementation study	Demonstrated clinician-adoption and follow-up-action gaps after AI-ECG deployment	Clinical implementation
Liu et al, 2024[17]	Asymptomatic LV dysfunction	Opportunistic AI-ECG screening strategy	Taiwanese outpatient cost-effectiveness model	Health-economic evaluation	Reported a negative ICER of -$7,439 for 65-year-old patients and cost-saving results across age groups in sensitivity analyses	Economic evaluation
Jacobs et al, 2024[18]	Newly abnormal LVEF after anthracycline therapy	AI-ECG screening for newly abnormal LVEF	Cardio-oncology population after anthracycline-based therapy	Screening/surveillance study	Supported AI-ECG as a surveillance tool for detecting newly abnormal LVEF after anthracycline exposure	Surveillance extension
Cohen-Shelly et al, 2021[19]	Aortic stenosis	12-lead ECG; deep learning	ECG-echocardiography cohort	Retrospective validation	AUC 0.85 for moderate-to-severe AS; AI-positive /echo-negative patients had increased future AS risk	Valvular model validation
Kwon et al, 2020[20]	Aortic stenosis	Deep learning ECG model	ECG-echocardiography dataset	Internal and external validation	AUC 0.884 internally and 0.861 externally	Valvular model validation
Elias et al, 2022; ValveNet[21]	Left-sided valvular heart disease	Deep learning ECG model	77,163 patients	Retrospective validation	AUROC 0.88 for AS, 0.77 for AR, and 0.83 for MR	Multi-lesion valvular validation
Vaid et al, 2023[22]	MR and AS	Deep learning ECG model	617,338 ECG-echocardiogram pairs from five hospitals	Multicenter retrospective validation	External-validation AUROC 0.81 for MR and 0.86 for AS	Multicenter valvular validation
Segar et al, 2026[23]	Aortic stenosis progression	AI-ECG model with longitudinal trajectory analysis	Community screening and pre-TAVR cohorts	External validation and longitudinal trajectory analysis	AI-positive discordance and longitudinal AI-ECG trajectories were associated with subsequent AS progression and AS-related events	Longitudinal validation
Liang et al, 2025[24]	Regurgitant valvular heart disease	AI-ECG models for diagnosis and future prediction of rVHD	International cohorts from ethnically distinct populations	International validation study	Developed and validated AI-ECG models to diagnose and predict future moderate or severe regurgitant valvular heart disease	Valvular prediction/ enrichment

Tsai et al, 2026[25]	Diastolic dysfunction after TAVR	AI-ECG-derived diastolic dysfunction grading and trajectory	Patients undergoing TAVR	Prognostic trajectory study	AI-ECG-derived diastolic dysfunction grade and trajectory provided dynamic prognostic stratification after TAVR	Post-intervention surveillance/trajectory analysis
Ko et al, 2020[26]	Hypertrophic cardiomyopathy	12-lead ECG; convolutional neural network	Mayo Clinic ECG cohort	Retrospective model development and validation	AUC 0.96; sensitivity 87%; specificity 90%	HCM model development
Goto et al, 2022[27]	Hypertrophic cardiomyopathy	Federated ECG and echocardiography models	Three academic centers in the United States and Japan	Multinational federated validation	Improved cross-site generalizability compared with single-institution models	Transportability-focused validation
Siontis et al, 2024[28]	Hypertrophic cardiomyopathy	AI-ECG model	External cohorts from Bern, Oxford, and Seoul	External validation	Confirmed good overall performance while showing site-level variability	International external validation
Sangha et al, 2025[29]	Hypertrophic cardiomyopathy	ECG image-based deep learning	Internal test set plus MIMIC-IV, AUMC, and UK Biobank external datasets	External validation	AUROC 0.95 internally and 0.94, 0.92, and 0.91 across external datasets	Image-based external validation
Goto et al, 2021[30]	Cardiac amyloidosis	AI-enabled ECG and echocardiography pipeline	Multicenter cohorts	Multicenter validation and deployment simulation	ECG C-statistics 0.85–0.91; ECG prescreening improved downstream echocardiographic PPV from 33% to 74–77% at 67% recall	Stepwise enrichment validation
Kwon et al, 2020[31]	Pulmonary hypertension	Deep learning ECG model	ECG cohort with PH labels	Model development and validation	Provided early evidence that ECG-based AI can predict pulmonary hypertension	Early PH model validation
Dubrock et al, 2024[32]	Pulmonary hypertension	AI-ECG algorithm	ECG-based PH detection cohort	Algorithm validation	Added validation evidence for AI-ECG detection of pulmonary hypertension	Early PH validation
Aras et al, 2023[33]	Pulmonary hypertension	Deep learning ECG model	ECG cohort	Retrospective validation	Supported feasibility of deep-learning ECG detection of pulmonary hypertension	Early PH validation
Ulloa-Cerna et al, 2022; rECHOmmend[9]	Composite SHD detectable by echocardiography	ECG-based machine learning using age, sex, and ECG traces	2.2 million ECGs from 484,765 adults	Retrospective development and validation	Composite AUROC 0.91; PPV 42% at 90% sensitivity	Composite referral model validation
Poterucha et al, 2025; EchoNext[3]	Composite SHD	Deep learning on ECG-echocardiography pairs	1.2 million ECG-echocardiogram pairs from 230,018 patients across eight hospitals	Multicenter validation plus prospective/silent-deployment components	Demonstrated broad SHD detection, referral enrichment, and the safety-net versus gatekeeper deployment framing	Multicenter translational validation
Dhingra et al, 2025; PRESENT-SHD[34]	Composite SHD	ECG image-based ensemble deep learning	Four U.S. hospitals and the prospective ELSA-Brasil cohort	External and prospective population-based validation	AUROC 0.853 in ELSA-Brasil; sensitivity 88%; specificity 62%; positive screen associated with 2- to 4-fold higher incident SHD/HF risk	Image-based external/prospective validation
Alexandrino et al, 2026[35]	Composite SHD workflow	AI-ECG followed by point-of-care cardiac ultrasound	486-patient pragmatic cohort; 286 patients with formal echocardiography reference standard	Proof-of-concept workflow study	AI-ECG PPV 32% and NPV 94%; adding POCUS increased PPV to 64%, maintained NPV at 93%, and improved accuracy from 67% to 88%	Stepwise confirmation feasibility

Note: This table corresponds to the 26 core evidence publications counted in Figure 1. Background epidemiology, treatment-context references, narrative reviews, editorials, methodological papers, general implementation/fairness/ethics literature, and adjacent non-AI-ECG studies cited for contextual discussion were not counted in this evidence set. Data size, model architecture, and performance metrics are reported when applicable; implementation, health-economic, and surveillance publications are summarized by their evaluation setting and pathway-relevant result.

Abbreviations: AI-ECG, artificial intelligence-enabled electrocardiography; AR, aortic regurgitation; AS, aortic stenosis; AUC/AUROC, area under the receiver operating characteristic curve; CMR, cardiac magnetic resonance; EF, ejection fraction; ECG, electrocardiogram; HCM, hypertrophic cardiomyopathy; ICER, incremental cost-effectiveness ratio; LV, left ventricular; LVEF, left ventricular ejection fraction; MR, mitral regurgitation; NPV, negative predictive value; PH, pulmonary hypertension; POCUS, point-of-care ultrasound; PPV, positive predictive value; rVHD, regurgitant valvular heart disease; SHD, structural heart disease; TAVR, transcatheter aortic valve replacement.

Table 2 Evidence Maturity and Deployment Implications of AI-ECG Across Structural Heart Disease Subtypes and Pathways

SHD Subtype or Pathway	Representative Evidence	Current Evidence Maturity	Safety-net Readiness	Gatekeeper Readiness	Main Evidence gaps	Practical Interpretation
Reduced LVEF/ LV dysfunction	Attia et al; EAGLE; inpatient, implementation, economic, and cardio-oncology surveillance studies[12–14,16–18]	Highest among current AI-ECG applications; includes retrospective validation, pragmatic randomized evidence, implementation work, and early economic evaluation	High	Low; no validated pathway for safe echocardiography deferral	Imaging-deferral thresholds; long-term clinical outcomes; follow-up adherence; multisystem implementation	Best-supported near-term use case; suitable as a safety-net trigger for confirmatory echocardiography rather than as a rule-out gatekeeper
Valvular heart disease	AS, MR/AR, rVHD, and post-TAVR trajectory studies[19–25]	Moderate; multiple retrospective and external validations, especially for AS, with emerging longitudinal and surveillance evidence	Moderate	Low; prospective deferral evidence is lacking	Prospective pathway trials; lesion-severity calibration; downstream echocardiography burden; actionable thresholds; subgroup performance	Promising referral-enrichment domain, especially for AS and left-sided valve disease, but not yet a routine clinical triage pathway
Hypertrophic cardiomyopathy	Ko et al; federated-learning HCM models; international external validation; ECG-image HCM model[26–29]	Moderate for detection; stronger transportability evidence than many rare-disease applications, but limited screening-pathway evidence	Low to moderate in targeted high-risk contexts	Very low; low-prevalence screening and downstream confirmation remain unresolved	Low-prevalence PPV; site-level variability; genetic/CMR confirmation pathways; prospective screening impact; downstream action protocols	Strong signal-detection field, but pathway-incomplete compared with low-LVEF AI-ECG
Cardiac amyloidosis	ECG-echocardiography AI pipeline with deployment simulation[30]	Early to moderate in enriched cohorts; limited direct evidence for broad population screening	Low in broad screening; potentially moderate in selected high-risk pathways	Very low; real-world PPV and confirmatory pathways remain limiting	Low disease prevalence; real-world PPV; external validation; confirmatory testing strategy; integration with amyloidosis diagnostic pathways	Useful as suspicion enrichment in selected patients; not appropriate as broad screening or gatekeeping evidence
Pulmonary hypertension/ right-heart dysfunction	PH AI-ECG studies and right-sided components in composite SHD models[3,31–33]	Early; still mainly retrospective signal-detection evidence with less standardized downstream pathways than left-sided SHD	Low	Very low; evidence remains mostly retrospective and pathway-incomplete	Reference-standard heterogeneity; PH subtype differences; prospective validation; referral workflow; clinical-action thresholds	Encouraging but exploratory; not yet a deployable echocardiography triage indication
Composite SHD models	rECHOmmend, EchoNext, and PRESENT-SHD[3,9,34]	Moderate and rapidly evolving; most clinically aligned with the question of whether a patient should receive echocardiography	Moderate to high for referral enrichment	Low; referral enrichment is supported, but rule-out use remains unvalidated	Randomized implementation; calibration; threshold selection; resource impact; clinician response; lesion-specific interpretability	Strong conceptual fit for safety-net screening and referral prioritization; gatekeeper use remains investigational

https://doi.org/10.2147/IJGM.S610298

Stepwise confirmation pathway	AI-ECG followed by point-of-care cardiac ultrasound[35]	Early proof-of-concept workflow evidence	Moderate as a second-step enrichment strategy	Low; evaluated for diagnostic-yield improvement rather than imaging deferral	Replication; operator variability; decentralized implementation; formal cost-effectiveness; integration with referral capacity	Attractive pathway for improving downstream diagnostic yield, but evidence remains preliminary
Longitudinal surveillance and risk trajectories	AS trajectories, rVHD prediction, cardio-oncology LVEF surveillance, and post-TAVR diastolic trajectory studies[18,23–25]	Emerging; extends AI-ECG from one-time detection toward surveillance and dynamic risk stratification	Potentially useful for follow-up prioritization	Very low; relevant to follow-up prioritization, not imaging replacement	Serial imaging validation; surveillance intervals; treatment-response linkage; clinical endpoint trials	Promising future direction, but not yet a substitute for established imaging surveillance protocols

Note: Evidence maturity was judged qualitatively according to the presence of retrospective model validation, external validation, prospective or randomized implementation evidence, health-economic evaluation, longitudinal/surveillance evidence, and direct evidence that AI-ECG changed downstream echocardiography referral or diagnostic yield. Safety-net readiness refers to use as an additive trigger or prioritization tool for confirmatory evaluation; gatekeeper readiness refers to use in deferring, withholding, or de-prioritizing echocardiography.

Abbreviations: AI-ECG, artificial intelligence-enabled electrocardiography; AR, aortic regurgitation; AS, aortic stenosis; CMR, cardiac magnetic resonance; HCM, hypertrophic cardiomyopathy; LVEF, left ventricular ejection fraction; MR, mitral regurgitation; PH, pulmonary hypertension; POCUS, point-of-care ultrasound; PPV, positive predictive value; rVHD, regurgitant valvular heart disease; SHD, structural heart disease; TAVR, transcatheter aortic valve replacement.

https://doi.org/10.2147/IJGM.S610298

Conceptual Framework: Safety-Net and Gatekeeper Use

The emerging literature suggests that AI-ECG should be understood not only as a diagnostic classifier, but as a workflow tool positioned upstream of echocardiography. EchoNext introduced the terms "safety net" and "gatekeeper", but these are better interpreted as intended-use orientations rather than fixed model categories.[3,13,36] This distinction provides the organizing framework for the evidence synthesis below. In a safety-net pathway, the central question is whether a positive AI-ECG result can improve detection by prompting otherwise-missed confirmatory evaluation. In a gatekeeper pathway, the central question is more demanding: whether a negative or low-risk AI-ECG result can safely support deferral or de-prioritization of echocardiography. The same algorithm may therefore be clinically acceptable as a trigger for additional imaging, yet still be insufficient as a basis for imaging deferral.

Safety-Net Use

In this review, safety-net deployment refers to an additive use case: AI-ECG is applied to ECGs already obtained in routine care to identify patients who might otherwise leave without echocardiographic evaluation. Its role is to enrich referral and support earlier recognition, not to overrule an existing indication for imaging. A positive AI-ECG result in this setting should therefore be interpreted as a structured trigger rather than a diagnosis. Depending on the phenotype and care setting, the downstream pathway may include clinician review of symptoms and prior imaging, confirmatory transthoracic echocardiography, focused point-of-care cardiac ultrasound as an intermediate enrichment step, or disease-specific referral after imaging confirmation. This is the deployment logic most clearly supported by current evidence. Randomized and pragmatic studies in low-ejection-fraction detection, including EAGLE and subsequent implementation work, have largely evaluated AI-ECG as a prompt for confirmatory echocardiography within routine workflows rather than as a tool to cancel imaging.[13,14,37,38] For safety-net use, model performance should therefore be judged by the balance among sensitivity, positive predictive value, referral burden, and follow-up completion, rather than by discrimination alone.

Gatekeeper Use

By contrast, gatekeeper deployment means that AI-ECG contributes to a decision to defer, de-prioritize, or withhold echocardiography in patients who might otherwise be imaged. This is a more demanding role because the consequence of error is no longer unnecessary testing alone, but potential false reassurance and missed structural heart disease. In such a pathway, a negative or low-risk AI-ECG result could only be considered in patients with sufficiently low pre-test probability and without independent clinical indications for echocardiography, such as heart-failure symptoms, concerning murmurs, known structural disease, abnormal biomarkers, or high-risk disease-specific contexts. Gatekeeper use requires more than discrimination: reliable calibration, very high sensitivity and negative predictive value in the intended population, prospective pathway-level validation, decision-curve evidence at the selected operating threshold, and explicit mechanisms for clinician override and follow-up.[36,39,40] On current evidence, that standard has not yet been established for broad structural heart disease screening.

Practical Framework

Accordingly, safety-net and gatekeeper use are treated not simply as labels, but as a framework for judging translational maturity. The key questions are whether AI-ECG meaningfully enriches echocardiography referral, whether that referral pathway has been tested prospectively within clinical workflow, and whether the workflow is sufficiently mature for deployment. By this standard, safety-net screening is the most mature near-term role, whereas gatekeeper triage remains aspirational pending stronger prospective evidence. Threshold choice, trade-offs, and workflow design are therefore not secondary details but determinants of deployability.

Evidence Synthesis Across SHD Domains

This section organizes the disease-specific evidence according to translational maturity rather than model performance alone. The core study-level evidence retained for qualitative synthesis is summarized in Table 1. Across SHD domains, we

evaluate each application using the same pathway-oriented dimensions: disease prevalence, validation depth, prospective pathway evidence, referral yield, downstream actionability, and readiness for safety-net versus gatekeeper deployment. The comparative maturity of each domain is summarized in Table 2. This structure allows reduced LVEF to serve as the benchmark indication, valvular and composite SHD models to be interpreted as intermediate referral-enrichment domains, and rarer or more heterogeneous phenotypes to be placed within earlier pathway-development stages.

Reduced LVEF: Benchmark Domain

The detection of reduced LVEF represents the most mature evidence domain for AI-ECG–guided echocardiography triage. The field was established by Attia et al in 2019, who demonstrated that a convolutional neural network could detect LVEF ≤35% with an AUC of 0.93 and a negative predictive value of 98.7%. In the same study, patients without baseline ventricular dysfunction but with a positive AI-ECG screen had a fourfold higher risk of developing future ventricular dysfunction than those with a negative screen.[12] This finding suggested that AI-ECG may identify not only prevalent ventricular dysfunction but also a higher-risk preclinical state.

The EAGLE trial (ECG AI-Guided Screening for Low Ejection Fraction) subsequently translated this algorithm into a pragmatic clinical intervention. In this trial, 120 primary care teams from 45 clinics or hospitals were cluster-randomized to either the intervention arm (access to AI results; 181 clinicians) or the control arm (usual care; 177 clinicians). ECGs were obtained as part of routine care from a total of 22,641 adults without prior heart failure. The primary outcome was a new diagnosis of low EF (≤50%) within 90 days of the ECG. The trial met this endpoint, increasing new diagnoses from 1.6% in the control arm to 2.1% in the intervention arm (OR 1.32, 95% CI 1.01–1.61; P = 0.007), equivalent to a 32% relative increase over usual care.[14]

Importantly, AI-enabled detection improved case finding without materially increasing overall echocardiography utilization, suggesting referral redirection rather than indiscriminate imaging expansion. Even so, follow-up imaging still occurred in only about half of AI-positive outpatients, and later inpatient pragmatic data likewise underscored a persistent gap between algorithmic flagging and downstream clinical action.[14,37]

Beyond derivation and primary-care randomization, the low-LVEF field has also advanced to additional prospective testing in distinct clinical settings. In the Nigeria obstetric trial, however, the significant diagnostic gain was driven by the digital-stethoscope pathway, whereas the 12-lead AI-ECG arm showed only a concordant but non-significant direction of effect; in Taiwan, a pragmatic inpatient randomized trial improved low-EF detection without increasing overall echocardiography utilization.[15,37] Taken together, the key question is not whether AI-ECG can detect low LVEF, but whether it has shown sufficient referral enrichment, prospective pathway validation, and workflow readiness to justify pre-echocardiographic deployment. On current evidence, low LVEF most clearly satisfies the first two criteria and has also moved furthest toward the third, making it the benchmark safety-net indication for near-term deployment. What remains unproven is not detection capability, but whether a negative or low-risk AI-ECG output can safely support echocardiography deferral.[3,36,39,40]

Valvular Heart Disease

Among non-LVEF applications, valvular heart disease—particularly aortic stenosis (AS)—has emerged as one of the more mature domains for AI-ECG screening. The prevalence of AS rises steeply with age, affecting approximately 1–2% of adults older than 65 years and about 12% of those older than 75 years, while definitive diagnosis and severity grading still rely on echocardiography.[41] AI-ECG algorithms for AS detection were developed independently by multiple groups. In the Mayo Clinic study, a convolutional neural network detected moderate-to-severe AS with an AUC of 0.85; importantly, patients with positive AI-ECG results but no moderate-to-severe AS at baseline had a 2.18-fold higher long-term risk of subsequently developing clinically significant AS.[19] Kwon et al independently reported AUCs of 0.884 in internal validation and 0.861 in external validation using a combined multilayer perceptron–convolutional neural network architecture.[20]

The Columbia-led ValveNet study extended AI-ECG screening beyond isolated AS to a broader left-sided VHD framework. In 77,163 patients, the model achieved area under receiver operating characteristic curve (AUROC) of 0.88 for AS, 0.77 for AR, and 0.83 for mitral regurgitation (MR), with similar overall accuracy in external validation.[21] Complementary

multicentre work from five Mount Sinai hospitals using 617,338 ECG–echocardiogram pairs likewise showed that ECG-based deep learning can detect both MR and AS, with external-validation AUROCs of 0.81 for MR and 0.86 for AS.[22]

Beyond discrimination, valvular AI-ECG—especially for aortic stenosis—also appears to identify a subset of discordant AI-positive/echo-negative patients who are at increased longitudinal risk. From a deployment perspective, however, the practical conclusion is narrower: valvular AI-ECG has moved beyond proof of concept and shows meaningful referral enrichment, but prospective pathway validation remains limited.[19,23] It is therefore best regarded as a translationally promising referral tool rather than a deployable pre-echocardiographic standard.

HCM and Cardiac Amyloidosis

HCM, with an estimated prevalence of approximately 1 in 500 in the general population, often remains clinically unrecognized and is associated with sudden cardiac death, including in younger individuals.[42] Ko et al at the Mayo Clinic developed the first AI-ECG model for HCM detection, achieving an AUC of 0.96 in an independent testing cohort, with 87% sensitivity and 90% specificity at the prespecified threshold. Performance remained strong in younger patients, in those with ECG-defined left ventricular hypertrophy, and even among patients with a normal ECG.[26]

A critical advance came with the application of federated learning to HCM detection. In a multinational Circulation study, ECG and echocardiography models trained across three academic centers in the United States and Japan showed substantially better cross-site generalizability than single-institution models and improved discrimination of HCM from other causes of hypertrophy.[27] External validation in Bern, Oxford, and Seoul subsequently confirmed good overall performance of the Mayo ECG model, but with clear site-level variability rather than uniform transportability across cohorts.[28] More recently, Sangha et al reported in a peer-reviewed study that a deep-learning model operating on ECG images—not raw waveform files—detected HCM with AUROCs of 0.95 in the internal test set and 0.94, 0.92, and 0.91 across three external datasets, supporting a potentially more deployable image-based screening route.[29]

Cardiac amyloidosis represents a distinct screening scenario because its low prevalence makes real-world positive predictive value (PPV) intrinsically challenging. Prior AI-enabled electrocardiography studies have reported encouraging discrimination for cardiac amyloidosis in enriched cohorts, suggesting that ECG-based screening may help enrich downstream diagnostic suspicion. However, the available evidence remains constrained by case-enriched designs, limited external validation, and the low-prevalence challenge intrinsic to real-world deployment.[30,43] The reported PPV of 0.86 arose from that matched case-control test set and should not be presented as a population-screening PPV.[30,43]

Goto et al subsequently reported a complementary ECG–echocardiography pipeline with C-statistics of 0.85–0.91 for ECG and 0.89–1.00 for echocardiography across multicenter cohorts. In deployment simulation, ECG prescreening improved the positive predictive value of downstream echocardiographic screening from 33% to 74–77% at 67% recall.[30]

For cardiac amyloidosis, the main translational challenge is not signal absence but low-prevalence deployment, because real-world positive predictive value will be substantially lower than performance observed in enriched case-control datasets. Existing studies suggest AI-ECG can enrich suspicion and improve the yield of downstream echocardiographic workup, but external validation remains limited and the evidence is not yet sufficient for routine population-level pre-echocardiographic triage. Taken together, HCM and CA should therefore be regarded as clinically important but pathway-incomplete domains rather than deployable screening indications comparable to low-LVEF AI-ECG.[26–30,43]

Pulmonary Hypertension and Right Heart Dysfunction

Evidence for AI-ECG detection of PH and right ventricular dysfunction remains comparatively early. Published studies are still predominantly retrospective, although reported discrimination has been encouraging, with AUROCs of approximately 0.86–0.92 across internal and external validation cohorts.[31–33] EchoNext also incorporated right-sided phenotypes into a broader composite SHD framework, suggesting that PH-related signals can be captured within referral-oriented screening.[3] Taken together, PH remains an early signal-detection domain rather than an actionable echocardiography-referral pathway: reported discrimination is encouraging, but the literature is still largely retrospective and the downstream pathway is less standardized than for left-sided SHD.

Composite SHD Models

A critical conceptual advance has been the development of models that simultaneously screen for multiple SHD entities, generating a single output — the estimated probability that a patient harbors any form of echocardiography-detectable structural disease. This approach better mirrors the clinical question of whether to order an echocardiogram.[9]

The rECHOmmend platform, developed by Ulloa-Cerna et al at Geisinger, used 2.2 million ECGs from 484,765 adults to train a composite model predicting any of seven echocardiography-confirmed conditions. The composite rECHOmmend model used age, sex, and ECG traces and achieved an AUROC of 0.91 with a positive predictive value of 42% at 90% sensitivity, outperforming the corresponding single-disease models in practical referral yield.[9] Individual disease models had AUROCs of 0.86–0.93 but substantially lower PPVs (1–31%), demonstrating that the composite approach improved practical utility.

EchoNext extended composite SHD screening to a multicenter, demographically diverse eight-hospital cohort and showed that broad referral-oriented detection can generalize across academic and community settings.[3] Prospective pilot data likewise showed a graded prevalence of previously unrecognized SHD across risk strata, supporting referral enrichment rather than binary rule-out.[3]

Dhingra et al developed PRESENT-SHD, an ensemble XGBoost model trained on ECG images rather than raw waveforms; the model was validated across four U.S. hospitals and the prospective ELSA-Brasil cohort, and maintained performance across novel ECG formats and smartphone photographs of monitors or paper printouts.[34] Composite models may be the most clinically aligned with the real-world referral question because they improve enrichment more than many lesion-specific models. However, current evidence still supports referral prioritization rather than imaging deferral, and the main trade-off remains reduced lesion-specific interpretability.

Overall, the evidence base is clinically promising but uneven in translational maturity. Reduced LVEF remains the benchmark indication because it has progressed from retrospective validation to pragmatic randomized implementation, follow-up adoption studies, and early economic evaluation. Valvular disease and composite SHD models represent the leading referral-enrichment domains, with encouraging retrospective, external-validation, and emerging longitudinal evidence but limited prospective pathway testing. HCM, cardiac amyloidosis, and pulmonary hypertension retain strong biological and clinical interest, yet their low prevalence, phenotypic heterogeneity, and less standardized downstream pathways limit broad screening readiness. This staged interpretation links the disease-specific evidence to the implementation questions addressed below: how positive results should trigger follow-up, how thresholds should be selected, and why gatekeeper use remains more demanding than safety-net deployment.

From Evidence to Implementation

The preceding evidence synthesis shows that AI-ECG applications differ substantially in maturity across SHD domains. The next question is how model outputs should be converted into clinical action. For safety-net use, implementation depends on whether positive results can increase diagnostic yield without overwhelming echocardiography capacity. For gatekeeper use, implementation depends on whether negative or low-risk outputs can support imaging deferral under validated thresholds and clinical safeguards. The following subsections therefore integrate false-positive interpretation, stepwise confirmation, threshold selection, workflow integration, prospective pathway evidence, and economic evaluation as linked implementation questions rather than separate technical topics.

Discordant Positive Results

In the best-studied AI-ECG screening settings—low LVEF and aortic stenosis—patients who screen positive on AI-ECG despite negative baseline echocardiography do not appear to be a low-risk group during follow-up. In the original Attia et al study, patients with false-positive low-EF screens had a fourfold higher risk of developing future ventricular dysfunction (hazard ratio 4.1; 95% CI 3.3–5.0).[12] Analogous findings have been reported in aortic stenosis: patients with positive AI-ECG results but no moderate-to-severe AS at baseline had a higher long-term risk of subsequently developing moderate-to-severe AS, and recent community validation likewise showed a fourfold higher risk of future AS hospitalization among false-positive screens.[19,23] More broadly, recent AI-ECG work in valvular disease suggests

that ECG-derived risk signals may reflect subclinical chamber remodelling and future progression risk, although this evidence is less mature than that for low LVEF and aortic stenosis.[24]

These observations have important implications for clinical pathway design. They are also consistent with newer AI-ECG studies suggesting that the ECG signal may capture future heart-failure risk beyond cross-sectional disease detection.[44,45] Rather than treating every discordant AI-positive/echo-negative result as a biologically meaningless false alarm, it is plausible that a subset of these results reflects preclinical disease, label-timing mismatch, or broader cardiovascular risk capture rather than pure algorithmic error; however, this interpretation should not obscure the fact that some discordant results will still represent clinically non-actionable false positives. At present, AI-ECG is better interpreted as a tool for risk enrichment and follow-up prioritization in selected screening settings than as a stand-alone mandate for structured surveillance or immediate pathway escalation.[12,19,44,45]

Stepwise Confirmation

A central challenge for scalable AI-ECG screening is that even referral-oriented composite models yield only moderate positive predictive value, so a substantial proportion of AI-positive patients will still proceed to confirmatory imaging. In rECHOmmend, the composite ECG model achieved a positive predictive value of 42% at 90% sensitivity, underscoring the persistent downstream burden of formal echocardiography despite improved referral enrichment.[9] Adjacent AI-echocardiography studies show that AI-assisted ultrasound can enable novice acquisition and focused assessment with encouraging diagnostic performance.[46,47] These adjacent ultrasound studies support technical feasibility of a second-step imaging filter, but they should not be conflated with direct validation of AI-ECG-triggered SHD triage pathways. Against this background, Alexandrino et al evaluated a two-tier workflow in which AI-ECG was followed by focused point-of-care cardiac ultrasound, with formal echocardiography as the reference standard. Among 286 patients with available reference echocardiography, the combined strategy increased positive predictive value from 32% to 64% while maintaining a negative predictive value of 93% and improving overall diagnostic accuracy from 67% to 88%, supporting referral enrichment rather than definitive rule-out. Applied to the full pragmatic cohort, this workflow yielded a number-needed-to-screen of 8 to identify 1 patient requiring formal echocardiography.[35]

Even so, this pathway should still be interpreted as early workflow-feasibility evidence rather than implementation-ready evidence. Direct evidence for AI-ECG-gated confirmation of multi-disease structural heart disease currently rests on a single published proof-of-concept study.[35] More broadly, recent echocardiography literature suggests that AI may support task shifting across image acquisition, analysis, and interpretation, but wider deployment still depends on careful validation, workflow integration, regulatory oversight, and clinically governed implementation.[48] Thus, the conceptual appeal of an AI-ECG-to-POCUS cascade is stronger than its current pathway-level evidence base, and its generalizability to decentralized routine screening remains unproven.

Thresholds and Decision Curves

The clinical utility of AI-ECG screening is not determined by discrimination alone. For a pre-echocardiographic triage tool, the critical question is whether a given probability threshold produces acceptable trade-offs among missed disease, unnecessary echocardiography, and downstream clinical benefit in the specific care setting. Decision curve analysis and net benefit are therefore more directly relevant to deployment than reporting a single AUROC, because they evaluate threshold-dependent clinical consequences rather than discrimination alone. Calibration is equally indispensable: a model may retain a high AUROC yet still provide unreliable absolute risk estimates and thereby misdirect referral decisions. NRI and IDI, if reported, should be treated as supplementary measures of relative model improvement rather than substitutes for calibration and decision-analytic assessment.[49,50]

Threshold selection should therefore be setting-specific and intended-use-specific rather than uniform. For safety-net deployment, thresholds should generally prioritize sensitivity and case-finding while maintaining an acceptable positive predictive value and manageable downstream echocardiography burden. In this scenario, false positives primarily affect confirmatory testing burden and downstream follow-up, and their acceptability depends on referral capacity, patient burden, and the extent to which discordant positive results represent future-risk enrichment. For gatekeeper deployment,

the threshold problem is different: the operating point must minimize false negatives, preserve very high negative predictive value in the local population, and show acceptable net benefit before imaging deferral can be considered. In emergency, inpatient, symptomatic, or high-risk settings, threshold choice should place greater weight on avoiding false reassurance. Future studies should therefore routinely report calibration plots, decision curves, net benefit, subgroup performance, and the rationale for context-specific operating thresholds, rather than treating sensitivity and specificity at a single cutoff as sufficient evidence of deployability.[50,51]

Workflow and the Action Gap

The transition from retrospective algorithm validation to clinical deployment requires that AI-ECG outputs be incorporated into EHR-based clinical workflows rather than remain stand-alone analytic results. This is not merely an informatics issue. When AI-ECG is used as a safety-net trigger, workflow friction mainly reduces follow-up efficiency; when it is proposed as a gatekeeper, the same friction may increase the risk of over-trusting low-risk outputs and delaying structural disease diagnosis.

In EAGLE, intervention clinicians were given access to AI-ECG results within routine primary care workflows, and these notifications were intended to prompt confirmatory echocardiography in patients flagged as potentially having low LVEF.[16,36] At this stage, the key issue is not merely whether AI-ECG can be displayed in the EHR, but whether that integration changes follow-up behavior without creating new workflow friction. Multicenter effectiveness data for EHR-integrated deployment remain limited, so claims about scalability, clinician adoption, and diagnostic yield should remain measured.[37]

Published low-LVEF AI-ECG implementation studies primarily support two practical workflow components—routine presentation of AI-ECG results and notification-based CDS—while other deployment architectures remain comparatively under-studied.[14,16] Population-management approaches remain conceptually plausible but under-studied in the AI-ECG literature. More broadly, CDS implementation research suggests that alert burden, clinical relevance, usability, and workflow disruption materially influence clinician uptake and downstream use.[52,53]

Even when AI-ECG models demonstrate excellent discrimination, their clinical impact remains contingent on clinician uptake and downstream testing. In EAGLE, only 49.6% of intervention-arm patients with a positive AI-ECG result underwent echocardiography, indicating that follow-up imaging still did not occur in roughly half of AI-positive outpatients even when the result was visible to clinicians.[14] In a temporally distinct silent-deployment cohort of EchoNext, 45% of patients labeled high risk did not undergo follow-up echocardiography as part of routine clinical care, highlighting a substantial action gap between model-based risk identification and downstream imaging.[3] These observations support prospective evaluation of whether active notification strategies can improve conversion of AI-ECG risk stratification into diagnostic follow-up in real-world screening pathways.

Alert fatigue and poor workflow integration are established barriers to CDS uptake. Systematic reviews of CDS implementation show that alert burden, clinical relevance and actionability, usability, and disruption of routine workflow materially influence clinician engagement and downstream use.[52,53] Implementation-oriented AI-ECG research has begun to use established implementation-science frameworks—particularly CFIR and ERIC—to identify barriers, facilitators, and deployment strategies; however, prospective framework-guided evaluations embedded within live AI-ECG rollout remain limited.[54] These implementation barriers are not peripheral; they are one reason why the safety-net model is currently more realistic than gatekeeper deployment in routine care, because a system that still struggles to convert positive AI alerts into timely follow-up is not yet ready to rely on negative AI-ECG outputs to support imaging deferral.

Prospective Pathway Evidence

The main contribution of EAGLE was not to re-establish that AI-ECG can detect low LVEF, which had already been shown in earlier derivation and validation studies, but to elevate the evidentiary level from model-performance literature to pragmatic care-pathway evidence.[14] By embedding AI-ECG outputs into routine primary-care workflows and evaluating downstream diagnosis through cluster randomization, EAGLE showed that an already validated signal could be operationalized as a real-world referral-enrichment strategy rather than remaining a stand-alone retrospective classifier.[14,55] The next evidentiary step is not another favorable single-system study but multisystem implementation

testing. AIM ECG-AI, together with contextual trials in Taiwan and Nigeria, extends the field toward generalizability across health systems and care environments.[15,37,38]

At this stage, generalizability is a more important evidentiary question than further single-system success. The AIM ECG-AI program matters because it is designed to move the field from single-system translation toward multicenter implementation testing, where heterogeneity in ECG acquisition, EHR integration, clinician response, and patient mix becomes part of the evidence rather than background noise.[38]

Related pragmatic trials in Taiwan and Nigeria are important less for re-establishing signal detection than for testing whether AI-assisted screening remains operationally workable across different care settings and resource environments.[15,37] Their value is therefore primarily contextual: Taiwan extends prospective implementation evidence into inpatient care, whereas Nigeria illustrates both the promise and the limits of AI-assisted cardiomyopathy screening in a low-resource obstetric setting, with the strongest randomized signal arising from the digital-stethoscope pathway rather than stand-alone 12-lead AI-ECG. Taken together, these studies suggest that the field is progressing from retrospective accuracy claims toward setting-specific implementation evidence, although durable multicenter effectiveness data and validated routine-deployment protocols remain limited.

For composite SHD models, the evidence base remains earlier than that for low-LVEF AI-ECG but is clearly moving beyond single-center retrospective waveform studies. In PRESENT-SHD, an image-based composite AI-ECG model validated across four U.S. hospitals also generalized to the prospective population-based ELSA-Brasil cohort, achieving an AUROC of 0.853 with 88% sensitivity and 62% specificity; a positive screen was further associated with a 2- to 4-fold higher risk of incident SHD or heart failure across external cohorts.[34] Earlier work from rECHOmmend likewise showed that combining multiple echocardiography-detectable conditions into a single referral-oriented endpoint improved practical yield, with an AUROC of 0.91 and a positive predictive value of 42% at 90% sensitivity, thereby supporting echocardiography-referral enrichment rather than lesion-by-lesion screening.[9] Even so, unlike low-LVEF AI-ECG—which already has pragmatic randomized trial evidence—broad SHD screening is still supported mainly by retrospective multicenter validation and nonrandomized prospective or population-based cohorts, and randomized implementation data, hard downstream outcome trials, and validated gatekeeper protocols for safely deferring echocardiography remain unavailable.[14,34] Trials such as ARISE remain informative for the broader question of AI-enabled cardiovascular alerting, but their primary contribution is AI-ECG-guided mortality-risk intervention in hospitalized patients rather than SHD-to-echocardiography triage; they are therefore better viewed as adjacent implementation evidence rather than direct evidence for SHD triage.[56]

Health-Economic Evidence

Among current AI-ECG applications, the economic evidence is most developed for low-LVEF screening, and importantly it aligns with an additive safety-net use case rather than a validated gatekeeper strategy. In a Taiwanese prospective outpatient study of opportunistic AI-ECG screening for asymptomatic left ventricular dysfunction, the strategy yielded a negative incremental cost-effectiveness ratio (ICER) of -$7,439 for 65-year-old patients and remained cost-saving across age groups in sensitivity analyses.[17] This economic signal is clinically plausible because screening for asymptomatic left ventricular systolic dysfunction has renewed relevance as medical therapy for preventing or delaying heart-failure progression has advanced.[57]

Beyond this Taiwanese outpatient study, peer-reviewed economic evidence remains limited rather than broadly confirmatory across health systems. By contrast, the Taiwanese ARISE economic analysis evaluated an inpatient AI-ECG mortality-alert strategy and reported an ICER of $59,500 per death averted; it should therefore be regarded as adjacent evidence for AI-enabled cardiovascular alerting rather than direct corroboration of low-LVEF screening cost-effectiveness.[58]

For composite SHD screening, formal cost-effectiveness evidence remains sparse, and peer-reviewed end-to-end economic validation has not yet been established.[3] A central economic concern in broader SHD screening is downstream testing burden: referral-oriented AI-ECG models can improve enrichment, yet moderate positive predictive value may still translate into substantial confirmatory echocardiography use.[9] Focused cardiac ultrasound is therefore conceptually attractive as an intermediate confirmation step. High-quality studies show that AI-guided or AI-assisted focused

ultrasound can be acquired and interpreted with good diagnostic performance, even by novice users, supporting the feasibility of two-step triage pathways.[46,47] However, robust high-impact economic evaluations and prospective implementation data for integrated AI-ECG+POCUS pathways are still lacking, so cost-saving claims should remain hypothetical rather than established.[3,35] At present, the economic literature supports the plausibility of downstream referral enrichment, but not the stronger claim that broader SHD gatekeeping can safely reduce imaging at the pathway level.

Future economic evaluations should explicitly define the standard-of-care comparator and the downstream management triggered by each test result; otherwise, pathway-level comparisons between AI-ECG alone and stepwise strategies will be difficult to interpret. Care-pathway mapping, probabilistic sensitivity analysis, and transparent handling of comparator variability will therefore be especially important in future AI-ECG economic models.[59] From a policy perspective, this asymmetry in economic evidence further reinforces that low-LVEF safety-net screening is closer to deployment maturity than broader SHD gatekeeping.

Challenges and Future Directions

The remaining barriers to AI-ECG-guided SHD triage extend beyond algorithmic accuracy. They can be organized into five linked domains: evidence standards, calibration and thresholding, workflow accountability, equity and transportability, and lifecycle governance. These challenges differ by intended use. Safety-net deployment primarily requires reliable follow-up pathways and sufficient referral capacity, whereas gatekeeper deployment additionally requires safeguards against false reassurance, missed disease, and inappropriate imaging deferral. This section therefore synthesizes these challenges through the safety-net versus gatekeeper framework before outlining future directions for surveillance, multimodal integration, and population-level screening.

Core Challenges by Intended Use

The first challenge is evidentiary asymmetry. A safety-net model may be useful when positive AI-ECG results enrich echocardiography yield and lead to timely confirmatory evaluation. Its main failure modes are excess downstream testing, incomplete follow-up, and alert fatigue. By contrast, gatekeeper use carries a different clinical risk profile because false-negative or poorly calibrated low-risk outputs may delay diagnosis. This distinction explains why gatekeeper deployment requires stronger evidence than safety-net deployment, including reliable calibration, high negative predictive value in the intended population, decision-curve evidence at the selected operating threshold, and prospective pathway-level validation.[16,38,52]

The second challenge is threshold transportability. Predictive values and net benefit are shaped by disease prevalence, referral context, ECG acquisition pathway, and patient mix. A threshold that is acceptable for opportunistic safety-net screening in primary care may not be appropriate for symptomatic, inpatient, emergency, or high-risk specialty settings. For safety-net use, threshold selection should balance case-finding against downstream imaging burden. For gatekeeper use, the operating point must preserve a very low false-negative risk across the local population and clinically important subgroups.

The third challenge is workflow accountability. AI-ECG outputs do not improve care unless responsibility for action is clearly assigned. Safety-net deployment requires closed-loop processes for notification, clinician review, confirmatory imaging, and follow-up of positive results. Gatekeeper deployment requires additional safeguards, including clinician override, auditability, re-entry into imaging pathways when symptoms or risk status change, and monitoring of delayed diagnoses. These requirements connect clinical decision support design directly to patient safety.[54–56]

The fourth challenge is equity and last-mile reliability. Fairness is not only a model-performance issue; it is also shaped by ECG acquisition quality, care setting, access to confirmatory echocardiography, and linkage to disease-specific treatment. Subgroup performance, input-format robustness, and post-deployment drift monitoring are therefore essential, particularly if AI-ECG is used beyond the development environment.[60–63] This issue is especially important for gatekeeper use, where uneven performance could translate into unequal access to imaging.

The fifth challenge is lifecycle governance. AI-ECG should be treated as a clinical pathway intervention rather than a static diagnostic score. Institutions need explicit rules for intended use, notification, patient communication, threshold updates, performance monitoring, and accountability for follow-up. In safety-net deployment, governance focuses on diagnostic yield and referral completion. In gatekeeper deployment, governance must also document the safety of

imaging deferral over time.[18,38] This intended-use-based framework provides a practical way to align validation standards, implementation design, and oversight with the clinical consequences of each deployment scenario.

Equity and Last-Mile Implementation

The rationale for AI-ECG screening may be strongest in low- and middle-income settings, where heart failure and other echocardiography-detectable cardiac disorders are often recognized late and access to echocardiography remains constrained by cost, workforce, and infrastructure. In such settings, the appeal of AI-ECG lies not in an already mature external-validation literature, but in the possibility of adding a scalable, low-cost triage layer upstream of formal imaging. Recent global reviews of heart failure and digital cardiovascular tools have accordingly emphasized that the unmet need may be greatest in LMICs, provided that implementation is accompanied by appropriate clinical pathways and treatment linkage.[64,65]

At present, however, peer-reviewed evidence specifically validating stand-alone 12-lead AI-ECG for structural screening in resource-limited African practice remains sparse. The strongest prospective African data come from a pragmatic randomized trial in Nigeria, which showed that AI-assisted screening for peripartum cardiomyopathy can be operationalized in hospital-based care; importantly, as noted earlier, the diagnostic gain in this trial was primarily driven by the AI-enabled digital-stethoscope pathway rather than the stand-alone 12-lead AI-ECG.[15] Current African evidence therefore supports the feasibility of AI-assisted screening in facility-based workflows, but does not yet provide a generalizable external-validation evidence base for stand-alone AI-ECG across rural primary care, community screening, or heterogeneous low-resource delivery systems. The next evidentiary step is prospective external-validation work with prespecified operating thresholds and explicit reporting of ECG acquisition quality, calibration, device heterogeneity, test failures, referral completion, and linkage to evidence-based therapy.

Equitable AI-ECG deployment depends not only on broadly similar performance across demographic subgroups, but also on transportability across institutions, acquisition workflows, and care environments. EchoNext reported broadly consistent performance across racial and care-setting strata, and federated-learning work in HCM supports improved cross-site generalizability.[3,27] Even so, average subgroup performance is not sufficient: clinically relevant degradation may still emerge in specific age-race-sex strata, and fairness in medical AI is shaped by data acquisition and workflow as well as by model architecture.[66,67]

Equity should therefore be treated as both a model property and a deployment property. A model that performs well under curated conditions may still fail when ECG inputs are obtained through heterogeneous real-world pathways, including image-based formats and smartphone photographs, unless performance is validated in the intended deployment environment.[34] Future validation should therefore report performance not only by subgroup, but also by care setting and input pathway, and should incorporate post-deployment monitoring for fairness drift.[60,61] For the present deployment question, unresolved transportability across input pathways is another reason to favor safety-net use over gatekeeper reliance.

Regulation and Trust

The regulatory maturity of AI-ECG for SHD screening remains uneven. Low-LVEF detection is the most mature use case, having already progressed from model development to randomized trials and early implementation studies.[12,14,16,37,38] By contrast, composite SHD screening and rarer-disease applications such as cardiac amyloidosis remain at an earlier stage, with the literature still dominated by model development, retrospective validation, and limited prospective implementation evidence.[3,43] This distinction matters because the evidentiary bar should differ by intended use: a safety-net application that prompts additional echocardiography may be justified by strong enrichment and acceptable referral yield, whereas any gatekeeper use that withholds or defers imaging should meet a substantially higher validation threshold before it can support echocardiography deferral in routine care.[3]

Explainability and ethics should be treated as linked components of clinical trust rather than as separate afterthoughts. Interpretability tools may make model behavior more inspectable, but they do not by themselves resolve questions of accountability, acceptable error, or downstream clinical action in opportunistic screening.[36,62] When AI-ECG is applied to a routine ECG obtained for another purpose, the output functions as an incidental risk signal and may trigger cascades of follow-up testing with uncertain net benefit.[63] For broad SHD deployment, trust will therefore depend not only on

transparent intended use and interpretable outputs where feasible, but also on proportionate oversight and explicit institutional rules for notification, patient communication, and follow-up responsibility.[62]

Surveillance and Multimodal Extensions

The evidence reviewed in this paper has primarily addressed AI-ECG as a tool for initial disease detection. However, AI-ECG may also evolve from a one-time detection tool into a longitudinal surveillance biomarker in selected settings, with emerging evidence supporting longitudinal risk tracking in aortic stenosis[23] and surveillance for treatment-related left ventricular dysfunction in cardio-oncology.[18] Recent studies have shown that longitudinal AI-ECG trajectories in aortic stenosis can identify accelerated disease progression years before valve intervention,[23] and AI-enabled electrocardiography-derived diastolic dysfunction trajectories after transcatheter aortic valve replacement also provide dynamic prognostic stratification;[25] in cardio-oncology, AI-ECG has shown promise for detecting newly abnormal LVEF after anthracycline therapy.[18] AI-ECG surveillance therefore represents a plausible extension of the screening paradigm, but it still requires prospective validation against serial imaging and clinical endpoints.

ECG foundation models may improve transportability and label efficiency, particularly for lower-prevalence phenotypes, but their translational value will depend on whether they improve calibration, robustness, and pathway performance rather than benchmark metrics alone.[68] Multimodal integration is most likely to matter when it reduces referral uncertainty in clinically ambiguous cases or strengthens the safety of downstream triage decisions, not simply when it produces incremental retrospective gains in discrimination. At present, however, direct SHD-specific evidence for ECG-centred multimodal screening remains limited, and convincing prospective evidence that such approaches can safely support gatekeeper-type decision making is still sparse.[69,70]

Population Screening Programs

Translating AI-ECG from diagnostic studies to scalable screening programmes will require decisions that extend beyond discrimination metrics, including whom to screen, how often to screen, what probability thresholds should trigger echocardiography, and whether downstream referral capacity, quality assurance, and outcome monitoring can support implementation.[36] Published AI-ECG studies have so far focused predominantly on opportunistic, clinic-based, or referral-enrichment strategies rather than organized population screening programmes with predefined invitation, recall, quality assurance, and long-term outcome monitoring structures.[3,14] Accordingly, organized population screening should still be regarded as a research agenda rather than a deployment-ready strategy. The next phase should therefore move beyond model validation toward programme-level evaluation of diagnostic yield, downstream imaging burden, cost-effectiveness, patient-centred outcomes, and governance feasibility.[36,71]

Conclusions

AI applied to the standard 12-lead electrocardiogram is reshaping the ECG from a rhythm-focused test into a potential upstream tool for structural heart disease triage. The strongest evidence is currently in low-LVEF screening, where the field has already progressed beyond retrospective validation to pragmatic randomized evaluation and early economic analysis. Valvular disease and composite SHD models are the leading expansion domains, whereas HCM, cardiac amyloidosis, and pulmonary hypertension remain earlier, pathway-incomplete applications.

Across indications, the decisive issue is clinical deploy ability rather than discrimination alone. On current evidence, safety-net use is the most credible near-term role, whereas gatekeeper deployment still lacks the validation, calibration, and operational safeguards required for routine echocardiography deferral. Near-term priorities are multicenter implementation, threshold optimization, prospective evaluation of stepwise confirmation pathways, and careful validation of longitudinal surveillance use cases. If implemented with appropriate safeguards, AI-ECG may improve clinical care by identifying otherwise-missed SHD, prioritizing limited echocardiography capacity, and accelerating referral to disease-specific treatment pathways.

Highlights

- AI-ECG is most mature as a safety-net trigger for echocardiography.
- Low-LVEF detection is the benchmark with randomized implementation evidence.

- Valvular and composite SHD models are promising but not yet pathway-mature.
- Gatekeeper use to defer imaging still lacks adequate prospective validation.
- Clinical impact will depend on workflow integration, equity, and referral capacity.

Declaration of Generative AI Use

Authors declare no AI use during the preparation of this work.

Abbreviations

SHD, structural heart disease; AI, artificial intelligence; ECG, electrocardiogram; LVEF, left ventricular ejection fraction; AS, aortic stenosis; MR, mitral regurgitation; AUROC, area under receiver operating characteristic curve; HCM, hypertrophic cardiomyopathy; PPV, positive predictive value; PH, pulmonary hypertension; HER, electronic health record; CDS, clinical decision support; ICER, incremental cost-effectiveness ratio.

Ethics Statement

This study did not directly involve the testing of human and animal samples; ethical approval is not applicable.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no competing interests in this work.

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovasc Res*. 2023;118(17):3272–3287. doi:10.1093/cvr/cvac013
2. AbouEzzeddine OF, Davies DR, Scott CG, et al. Prevalence of transthyretin amyloid cardiomyopathy in heart failure with preserved ejection fraction. *JAMA Cardiol*. 2021;6(11):1267–1274. doi:10.1001/jamacardio.2021.3070
3. Poterucha TJ, Jing L, Ricart RP, et al. Detecting structural heart disease from electrocardiograms using AI. *Nature*. 2025;644(8075):221–230. doi:10.1038/s41586-025-09227-0
4. Wang TJ, Levy D, Benjamin EJ, Vasan RS. The epidemiology of asymptomatic left ventricular systolic dysfunction: implications for screening. *Ann Intern Med*. 2003;138(11):907–916. doi:10.7326/0003-4819-138-11-200306030-00012
5. Carabello BA, Paulus WJ. Aortic stenosis. *Lancet*. 2009;373:956–966. doi:10.1016/s0140-6736(09)60211-7
6. Witteles RM, Bokhari S, Damy T, et al. Screening for transthyretin amyloid cardiomyopathy in everyday practice. *JACC Heart Fail*. 2019;7(8):709–716. doi:10.1016/j.jchf.2019.04.010
7. Olivotto I, Oreziak A, Barriales-Villa R, et al. Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, Phase 3 trial. *Lancet*. 2020;396(10253):759–769. doi:10.1016/s0140-6736(20)31792-x
8. Maurer MS, Schwartz JH, Gundapaneni B, et al. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med*. 2018;379(11):1007–1016. doi:10.1056/NEJMoa1805689
9. Ulloa-Cerna AE, Jing L, Pfeifer JM, et al. rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. *Circulation*. 2022;146(1):36–47. doi:10.1161/circulationaha.121.057869
10. van de Leur RR, Blom LJ, Gavves E, et al. Automatic triage of 12-Lead ECGs using deep convolutional neural networks. *J Am Heart Assoc*. 2020;9(10):e015138. doi:10.1161/jaha.119.015138
11. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18:465–478. doi:10.1038/s41569-020-00503-2
12. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med*. 2019;25(1):70–74. doi:10.1038/s41591-018-0240-2
13. Antoniades C, Chan K. Scalable screening for structural heart disease: promises from artificial intelligence-electrocardiogram tools. *Euro Heart J Digital Health*. 2025;6(4):521–523. doi:10.1093/ehjdh/ztaf048

14. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. 2021;27(5):815–819. doi:10.1038/s41591-021-01335-4
15. Adedinsewo DA, Morales-Lara AC, Afolabi BB, et al. Artificial intelligence guided screening for cardiomyopathies in an obstetric population: a pragmatic randomized clinical trial. *Nat Med*. 2024;30(10):2897–2906. doi:10.1038/s41591-024-03243-9
16. Rushlow DR, Croghan IT, Inselman JW, et al. Clinician adoption of an artificial intelligence algorithm to detect left ventricular systolic dysfunction in primary care. *Mayo Clin Proc*. 2022;97(11):2076–2085. doi:10.1016/j.mayocp.2022.04.008
17. Liu W-T, Hsieh P-H, Lin C-S, et al. Opportunistic screening for asymptomatic left ventricular dysfunction with the use of electrocardiographic artificial intelligence: a cost-effectiveness approach. *Can J Cardiol*. 2024;40(7):1310–1321. doi:10.1016/j.cjca.2023.11.044
18. Jacobs JEJ, Greason G, Mangold KE, et al. Artificial intelligence electrocardiogram as a novel screening tool to detect a newly abnormal left ventricular ejection fraction after anthracycline-based cancer therapy. *Eur J Prev Cardiol*. 2024;31(5):560–566. doi:10.1093/eurjpc/zwad348
19. Cohen-Shelly M, Attia ZI, Friedman PA, et al. Electrocardiogram screening for aortic valve stenosis using artificial intelligence. *Eur Heart J*. 2021;42(30):2885–2896. doi:10.1093/eurheartj/ehab153
20. Kwon J-M, Lee SY, Jeon K-H, et al. Deep learning–based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc*. 2020;9(7):e014717. doi:10.1161/jaha.119.014717
21. Elias P, Poterucha TJ, Rajaram V, et al. Deep learning electrocardiographic analysis for detection of left-sided valvular heart disease. *J Am Coll Cardiol*. 2022;80(6):613–626. doi:10.1016/j.jacc.2022.05.029
22. Vaid A, Argulian E, Lerakis S, et al. Multi-center retrospective cohort study applying deep learning to electrocardiograms to identify left heart valvular dysfunction. *Commun Med*. 2023;3(1):24. doi:10.1038/s43856-023-00240-w
23. Segar MW, Lambeth KD, Postalian A, et al. Validation and longitudinal trajectory analysis of an AI-based ECG model for aortic stenosis: from community screening to pre-TAVR risk stratification. *Euro Heart J Digital Health*. 2026;7(2):ztag018. doi:10.1093/ehjdh/ztag018
24. Liang Y, Sau A, Zeidaabadi B, et al. Artificial intelligence-enhanced electrocardiography to predict regurgitant valvular heart diseases: an international study. *Eur Heart J*. 2025;46(44):4823–4837. doi:10.1093/eurheartj/ehaf448
25. Tsai C-M, Naser JA, Tsaban G, et al. Prognostic value of artificial intelligence–enabled electrocardiography–derived diastolic dysfunction grading and trajectory in patients undergoing transcatheter aortic valve replacement. *J Am Heart Assoc*. 2026;15(3):e046558. doi:10.1161/jaha.125.046558
26. Ko W-Y, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75(7):722–733. doi:10.1016/j.jacc.2019.12.030
27. Goto S, Solanki D, John JE, et al. Multinational federated learning approach to train ECG and echocardiogram models for hypertrophic cardiomyopathy detection. *Circulation*. 2022;146(10):755–769. doi:10.1161/circulationaha.121.058696
28. Siontis KC, Wieczorek MA, Maanja M, et al. Hypertrophic cardiomyopathy detection with artificial intelligence electrocardiography in international cohorts: an external validation study. *Euro Heart J Digital Health*. 2024;5(4):416–426. doi:10.1093/ehjdh/ztae029
29. Sangha V, Dhingra LS, Aminorroaya A, et al. Identification of hypertrophic cardiomyopathy on electrocardiographic images with deep learning. *Nat Cardiovasc Res*. 2025;4(8):991–1000. doi:10.1038/s44161-025-00685-3
30. Goto S, Mahara K, Beussink-Nelson L, et al. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nat Commun*. 2021;12(1):2726. doi:10.1038/s41467-021-22877-8
31. Kwon J-M, Kim K-H, Medina-Inojosa J, Jeon K-H, Park J, Oh B-H. Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography. *J Heart Lung Transplant*. 2020;39(8):805–814. doi:10.1016/j.healun.2020.04.009
32. DuBrock HM, Wagner TE, Carlson K, et al. An electrocardiogram-based AI algorithm for early detection of pulmonary hypertension. *Eur Respir J*. 2024;64(1):. doi:10.1183/13993003.00192-2024
33. Aras MA, Abreau S, Mills H, et al. Electrocardiogram detection of pulmonary hypertension using deep learning. *J Card Fail*. 2023;29 (7):1017–1028. doi:10.1016/j.cardfail.2022.12.016
34. Dhingra LS, Aminorroaya A, Sangha V, et al. Ensemble deep learning algorithm for structural heart disease screening using electrocardiographic images. *J Am Coll Cardiol*. 2025;85(12):1302–1313. doi:10.1016/j.jacc.2025.01.030
35. Alexandrino FB, Schlesinger R, Bird J, et al. Integrating AI-ECG and point-of-care cardiac ultrasound for screening structural heart disease: a proof-of-concept study. *Am Heart J*. 2026;294:107337. doi:10.1016/j.ahj.2025.107337
36. Lin CS, Liu WT, Chen YH, Lin SH, Lin C. Artificial intelligence-enabled electrocardiography from scientific research to clinical application. *EMBO Mol Med*. 2026;18:22–40. doi:10.1038/s44321-025-00351-y
37. Tsai D-J, Lin C, Liu W-T, et al. Artificial intelligence-assisted diagnosis and prognostication in low ejection fraction using electrocardiograms in inpatient department: a pragmatic randomized controlled trial. *BMC Med*. 2025;23(1):342. doi:10.1186/s12916-025-04190-z
38. Lopez-Jimenez F, Alger HM, Attia ZI, et al. A multicenter pragmatic implementation study of AI-ECG-based clinical decision support software to identify low LVEF: clinical trial design and methods. *Am Heart J Plus*. 2025;54:100528. doi:10.1016/j.ahjo.2025.100528
39. Oikonomou EK, Khera R. Artificial intelligence-enhanced patient evaluation: bridging art and science. *Eur Heart J*. 2024;45:3204–3218. doi:10.1093/eurheartj/ehae415
40. Palermi S, Vecchiato M, Ng FS, et al. Artificial intelligence and the electrocardiogram: a modern renaissance. *Eur J Intern Med*. 2025;140:106329. doi:10.1016/j.ejim.2025.04.036
41. Otto CM, Newby DE, Hillis GS. Calcific Aortic Stenosis: a Review. *JAMA*. 2024;332:2014–2026. doi:10.1001/jama.2024.16477
42. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242–255. doi:10.1016/s0140-6736(12)60397-3
43. Grogan M, Lopez-Jimenez F, Guthrie S, et al. Value of artificial intelligence for enhancing suspicion of cardiac amyloidosis using electrocardiography and echocardiography: a narrative review. *J Am Heart Assoc*. 2025;14(8):e036533. doi:10.1161/jaha.124.036533
44. Desai AS, Pandey A, Suratekar R, et al. Predicting heart failure from 12-lead ECGs using AI: a HEARTSHARE/AMP-HF pooled cohort analysis. *J Am Coll Cardiol*. 2026;87(8):990–1005. doi:10.1016/j.jacc.2025.10.065
45. Khurshid S, Friedman SF, Kany S, et al. Artificial intelligence-enabled ECG analysis to predict incident heart failure. *Circ Heart Fail*. 2026;19(4): e013927. doi:10.1161/circheartfailure.125.013927
46. Mor-Avi V, Khandheria B, Klempfner R, et al. Real-Time Artificial Intelligence–Based Guidance of Echocardiographic Imaging by Novices: image Quality and Suitability for Diagnostic Interpretation and Quantitative Analysis. *Circ Cardiovasc Imaging*. 2023;16(11):e015569. doi:10.1161/circimaging.123.015569

47. Motazedian P, Marbach JA, Prosperi-Porta G, et al. Diagnostic accuracy of point-of-care ultrasound with artificial intelligence-assisted assessment of left ventricular ejection fraction. *NPJ Digit Med*. 2023;6(1):201. doi:10.1038/s41746-023-00945-1
48. Myhre PL, Grenne B, Asch FM, et al. Artificial intelligence-enhanced echocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2026;23(3):164–182. doi:10.1038/s41569-025-01197-0
49. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–138. doi:10.1097/EDE.0b013e3181c30fb2
50. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA*. 2017;318(14):1377–1384. doi:10.1001/jama.2017.12126
51. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6. doi:10.1136/bmj.i6
52. Meunier P-Y, Raynaud C, Guimaraes E, Gueyffier F, Letrilliart L. Barriers and facilitators to the use of clinical decision support systems in primary care: a mixed-methods systematic review. *Ann Fam Med*. 2023;21(1):57–69. doi:10.1370/afm.2908
53. Van Dort BA, Zheng WY, Sundar V, Baysari MT. Optimizing clinical decision support alerts in electronic medical records: a systematic review of reported strategies adopted by hospitals. *J Am Med Inform Assoc*. 2021;28(1):177–183. doi:10.1093/jamia/ocaa279
54. Arends BKO, McCormick JM, van der Harst P, Heus P, van Es R. Barriers, facilitators and strategies for the implementation of artificial intelligence-based electrocardiogram interpretation: a mixed-methods study. *Eur J Clin Invest*. 2025;55(S1). doi:10.1111/eci.14387
55. Yao X, McCoy RG, Friedman PA, et al. ECG AI-Guided Screening for Low Ejection Fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. *Am Heart J*. 2020;219:31–36. doi:10.1016/j.ahj.2019.10.007
56. Lin C-S, Liu W-T, Tsai D-J, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. *Nat Med*. 2024;30(5):1461–1470. doi:10.1038/s41591-024-02961-4
57. Bjerkén LV, Rønborg SN, Jensen MT, Ørting SN, Nielsen OW. Artificial intelligence enabled ECG screening for left ventricular systolic dysfunction: a systematic review. *Heart Fail Rev*. 2022;28(2):419–430. doi:10.1007/s10741-022-10283-1
58. Hsieh P-H, Lin C, Lin C-S, et al. Economic analysis of an AI-enabled ECG alert system: impact on mortality outcomes from a pragmatic randomized trial. *NPJ Digit Med*. 2025;8(1):348. doi:10.1038/s41746-025-01735-7
59. Graziadio S, Gregg E, Allen AJ, et al. Is the comparator in your diagnostic cost-effectiveness model "standard of Care"? Recommendations from literature reviews and expert interviews on how to identify and operationalize it. *Value Health*. 2024;27(5):585–597. doi:10.1016/j.jval.2024.02.003
60. Zhou X, Li T, Hayama H, et al. Diagnosis of cardiac conditions from 12-lead electrocardiogram through natural language supervision. *NPJ Digit Med*. 2025;8(1):697. doi:10.1038/s41746-025-02074-3
61. Davis SE, Dorn C, Park DJ, Matheny ME. Emerging algorithmic bias: fairness drift as the next dimension of model maintenance and sustainability. *J Am Med Inform Assoc*. 2025;32(5):845–854. doi:10.1093/jamia/ocaf039
62. Lewin S, Chetty R, Ihdayhid AR, Dwivedi G. Ethical challenges and opportunities in applying artificial intelligence to cardiovascular medicine. *Can J Cardiol*. 2024;40(10):1897–1906. doi:10.1016/j.cjca.2024.06.029
63. Ganguli I, Simpkin AL, Lupo C, et al. Cascades of care after incidental findings in a US national survey of physicians. *JAMA Netw Open*. 2019;2(10):e1913325. doi:10.1001/jamanetworkopen.2019.13325
64. Mbanze I, Spracklen TF, Jessen N, Damasceno A, Sliwa K. Heart failure in low-income and middle-income countries. *Heart*. 2025;111(8):341–351. doi:10.1136/heartjnl-2024-324176
65. Myhre PL, Tromp J, Ouwerkerk W, et al. Digital tools in heart failure: addressing unmet needs. *Lancet Digit Health*. 2024;6(10):e755–e766. doi:10.1016/s2589-7500(24)00158-4
66. Kaur D, Hughes JW, Rogers AJ, et al. Race, sex, and age disparities in the performance of ECG deep learning models predicting heart failure. *Circ Heart Fail*. 2024;17(1):e010879. doi:10.1161/circheartfailure.123.010879
67. Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023;7(6):719–742. doi:10.1038/s41551-023-01056-8
68. Nolin-Lapalme A, Sowa A, Delfrate J, et al. Foundation models for electrocardiogram interpretation: clinical implications. *Eur Heart J*. 2026;47(18):2174–2186. doi:10.1093/eurheartj/ehaf1119
69. Turgut Ö, Müller P, Hager P, et al. Unlocking the diagnostic potential of electrocardiograms through information transfer from cardiac magnetic resonance imaging. *Med Image Anal*. 2025;101:103451. doi:10.1016/j.media.2024.103451
70. Kishikawa R, Kodera S, Setoguchi N, et al. An ensemble learning model for detection of pulmonary hypertension using electrocardiogram, chest X-ray, and brain natriuretic peptide. *Europ Heart J Digital Health*. 2025;6(2):209–217. doi:10.1093/ehjdh/ztae097
71. Motazedian P, Prosperi-Porta G, Hibbert B, et al. Cost-effectiveness of population screening for aortic stenosis. *Eur Heart J Qual Care Clin Outcomes*. 2025;11(4):378–387. doi:10.1093/ehjqcco/qcae043