




The diagnostic accuracy of artificial intelligence enhanced electrocardiography for the detection of cardiac dysfunction

Habib Shahzad^{1*} , Humna Sulaiman¹, Amna Roheel², Hasan N. Tahir³, Mursala Tahir⁴, Mahad Tariq¹, Hooria Sulaiman¹

¹Rawalpindi Medical University, Rawalpindi 46000, Pakistan

²Riphah International University, Islamabad 54000, Pakistan

³Department of Community medicine, College of Medicine, Shaqra University, Dawadami 11961, Saudi Arabia

⁴Department of Community Medicine, Jinnah Sindh Medical University, Karachi 75510, Pakistan

***Correspondence:** Habib Shahzad, Rawalpindi Medical University, Rawalpindi 46000, Pakistan. shehzadhabib923@gmail.com

Academic Editor: Teresa Tsang, University of British Columbia, Canada

Received: December 21, 2025 **Accepted:** January 26, 2026 **Published:** February 5, 2026

Cite this article: Shahzad H, Sulaiman H, Roheel A, Tahir HN, Tahir M, Tariq M, et al. The diagnostic accuracy of artificial intelligence enhanced electrocardiography for the detection of cardiac dysfunction. *Explor Cardiol.* 2026;4:101293. <https://doi.org/10.37349/ec.2026.101293>

Abstract

Background: Heart failure (HF) remains a growing global health problem, with nearly half of all cases attributed to HF with preserved ejection fraction (HFpEF) and its precursor, left ventricular diastolic dysfunction (LVDD). Although echocardiography is the diagnostic gold standard, its high cost and limited availability restrict its use for large-scale screening. In contrast, the electrocardiogram (ECG) is inexpensive and widely accessible. Recent advances in artificial intelligence (AI) have created opportunities to leverage ECG data for the early detection of cardiac dysfunction. The objective of this study was to systematically review and meta-analyze the diagnostic performance of AI-based ECG models for detecting cardiac dysfunction.

Methods: The QUADAS-2 tool was used to assess the risk of bias. Pooled sensitivity and specificity were estimated using a bivariate random-effects model, with heterogeneity quantified using the I^2 statistic. Pre-specified subgroup analyses were conducted according to clinical endpoint and AI model type.

Results: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, nine eligible studies evaluating AI algorithms applied to ECG data for the detection of HFpEF were identified. Considerable methodological and population heterogeneity was observed across studies. Risk of bias was generally low for reference standards, although concerns were noted in patient selection. The pooled specificity of AI-ECG models was high at 0.83 [95% confidence interval (CI): 0.74–0.89], while pooled sensitivity was 0.82 (95% CI: 0.70–0.90). Both estimates demonstrated extremely high heterogeneity ($I^2 > 96\%$). Subgroup analyses by endpoint and model type did not explain this variability.

Discussion: AI-enhanced ECG models show good diagnostic accuracy, specifically in ruling out cardiac dysfunction due to their high specificity. However, the high and unexplained heterogeneity across these studies limits the immediate generalizability of the results. Large, prospective validation studies across



diverse populations are essential before these models can be confidently adopted into routine clinical practice.

Keywords

heart failure with preserved ejection fraction, artificial intelligence, echocardiogram

Introduction

Heart failure (HF) is a life-threatening syndrome with rising global prevalence and is associated with significant morbidity, mortality, and healthcare expenditure [1–3]. Globally, HF affects an estimated 64 million individuals. Nearly half of all HF cases are now attributed to HF with preserved ejection fraction (HFpEF) [4, 5]. Moreover, with the aging of the population, the prevalence of HFpEF is expected to increase in the coming years, reflecting both increased life expectancy and the expanding proportion of older individuals in the general population [5]. The pathophysiological hallmark of HFpEF is left ventricular diastolic dysfunction (LVDD), which often develops gradually and may remain clinically silent until progression to overt HF [6]. Therefore, early diagnosis of HFpEF can play a key role in facilitating prompt initiation of drugs that reduce the burden of HF in this population [7, 8].

Currently, echocardiography serves as the gold standard for diagnosing LVDD and HFpEF, offering detailed structural and functional insights [9, 10]. However, it relies on costly equipment and specialized expertise, making it unsuitable for large-scale screening. Invasive methods, such as right heart catheterization, provide definitive confirmation of elevated filling pressures but are restricted to selected cases given their risks and costs [11]. These limitations contribute to delays in diagnosis, reducing opportunities for timely intervention.

The electrocardiogram (ECG), in contrast, is inexpensive, widely available, and noninvasive. While subtle electrical changes in HFpEF and LVDD may appear on ECGs, they often go unnoticed by conventional human interpretation. In the recent past, with the development and widespread application of artificial intelligence (AI) in medicine, we have entered a novel era of diagnosis of disease and management. The efficient processing ability of AI achieves specific goals and helps clinicians a lot in clinical decision-making. AI mainly includes machine learning (ML) and deep learning (DL), and DL consists of multiple processing layers and is capable of processing more complex data [12]. The application of AI has contributed to clinical practice in terms of aiding diagnosis [13] and early detection [14].

One area of interest that is relevant to our topic is the combination of AI with ECG for the detection of HF, which may overcome some difficulties in clinical diagnosis [15]. Several studies have specifically investigated their role in detecting HFpEF, LVDD, or raised filling pressures. However, these studies vary a lot in their design, target populations, and diagnostic endpoints. To synthesize this heterogeneous evidence, we performed a systematic review and meta-analysis to estimate the pooled diagnostic performance of AI-based ECG models for the detection of HFpEF, LVDD, and related cardiac dysfunction. HFpEF, LVDD, and elevated left ventricular filling pressures were considered related clinical manifestations along the same disease continuum. HFpEF represents the clinical syndrome, while LVDD and elevated filling pressures reflect earlier or mechanistic stages detectable by imaging or invasive assessment.

Materials and methods

Protocol and reporting

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [16].

Eligibility criteria and information sources

We included studies that (1) applied an AI or ML algorithm; (2) used ECG data (single-lead or 12-lead) as the input; (3) focused on detecting or predicting cardiac dysfunction related to impaired diastolic function,

including HFpEF, LVDD, or objectively confirmed elevated filling pressures; and (4) reported sufficient data to construct a 2×2 contingency table. We found a total of nine studies that were available for inclusion in this analysis.

Study selection and search strategy

PubMed, Cochrane, CTG, and Google Scholar databases were systematically searched to identify original literature that evaluated the diagnostic accuracy of AI algorithms using ECG data for people with HFpEF. The full search strategies for all databases are provided in [Supplementary materials](#).

Two review authors independently performed the literature search and title/abstract screening against predefined criteria, with full texts reviewed for eligibility, and the articles were excluded that were not associated with the research topic. Nine studies met the inclusion criteria and were incorporated into both qualitative and quantitative synthesis. Results of the literature search are shown in [Figure 1](#).

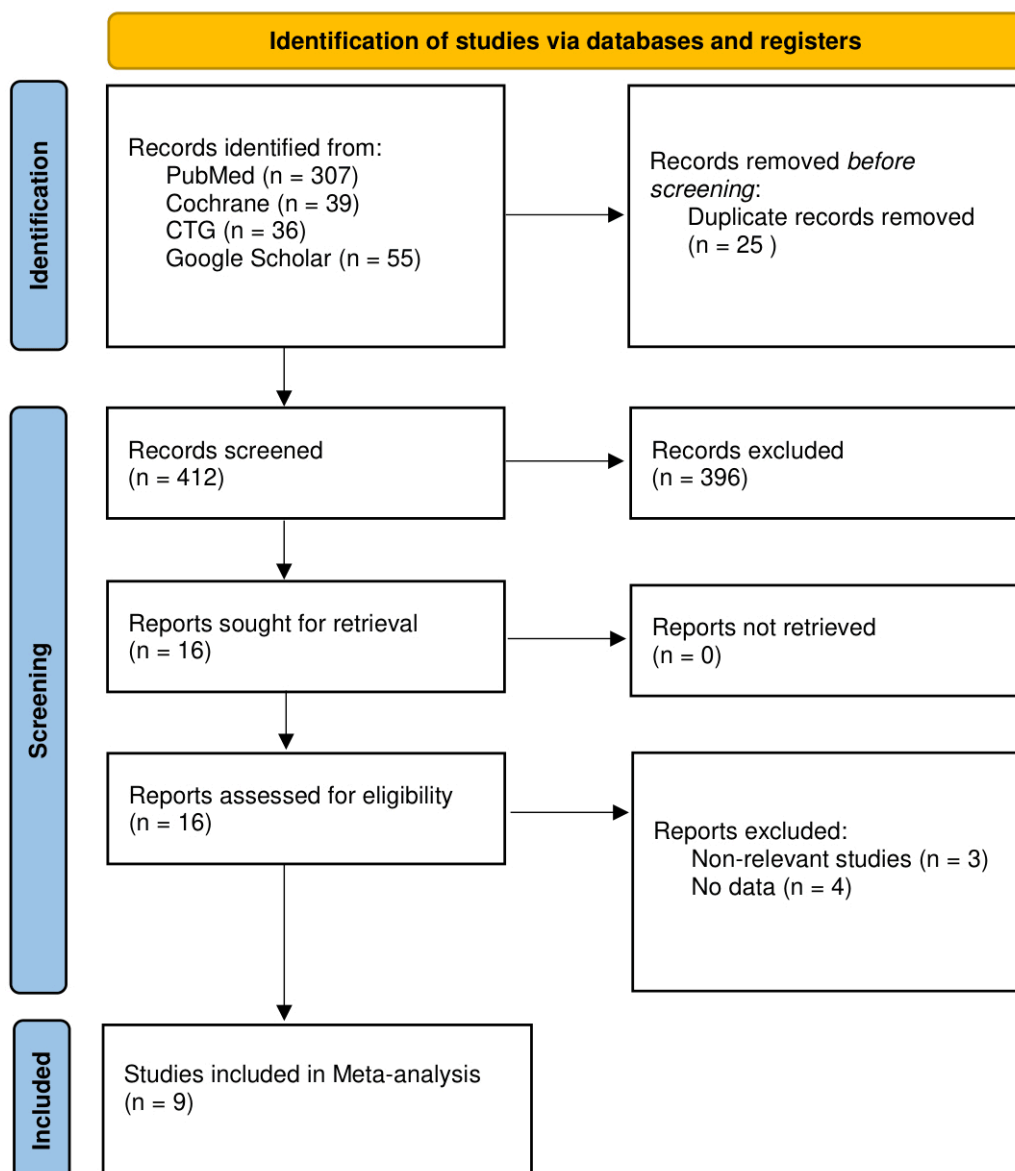


Figure 1. Flow diagram of the study selection process. Adapted from [16]. © Author(s) (or their employer(s)) 2019. CC BY. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Data extraction

For each study, we extracted the first author, publication year, study design, population characteristics, AI/ML model type, diagnostic endpoint, and reported performance metrics. For the meta-analysis, we

obtained or derived true positives, false positives, true negatives, and false negatives from data reported in the original publications, including confusion matrices, diagnostic accuracy tables, or reported sensitivity, specificity, and sample size. No AI models were trained or revalidated by the authors; all analyses were based exclusively on published study results.

Risk of bias assessment

Quality of studies was independently assessed using the QUADAS-2 tool [17], which evaluates four domains: Patient Selection, Index Test, Reference Standard, and Flow/Timing. Each was classified as low, high, or some risk of bias. Two reviewers independently performed the data extraction and quality assessment. Disagreements were resolved through discussion and independent assessment by another researcher to reach a consensus. A total of two disagreements (out of 2 assessments) were identified, all of which were resolved through consensus discussion.

Statistical analysis

Our primary outcomes were the pooled sensitivity and specificity of AI-ECG models. Sensitivity was calculated as $TP/(TP + FN)$, representing the proportion of true disease cases correctly identified. Specificity was calculated as $TN/(TN + FP)$, representing the proportion of non-diseased cases correctly classified. We used bivariate random-effects for meta-analysis, allowing joint estimation of sensitivity and specificity while accounting for their correlation. This approach produced a summary operating point and a hierarchical summary receiver operating characteristic (SROC) curve with 95% confidence intervals (CIs) and prediction intervals.

Heterogeneity was quantified using the I^2 statistic, with values above 75% considered substantial. To explore potential sources of heterogeneity, we conducted subgroup analyses by (1) clinical endpoint (HFpEF, LVDD, or increased filling pressure) and (2) AI model type (DL vs. classical ML). All analyses were performed in *R*.

HFpEF, LVDD, and elevated filling pressures were analyzed together as diagnostic endpoints reflecting diastolic dysfunction, given their shared pathophysiological basis and overlapping diagnostic criteria.

Results

Study selection and characteristics

The study selection process is depicted in the PRISMA flow diagram (Figure 1). Nine studies were ultimately included. They varied a lot in their designs, study populations, and methodological approaches (Table 1). A wide range of AI models were evaluated, including DL architectures such as convolutional neural networks (CNNs) and classical algorithms like Random Forest and XGBoost. ECG inputs ranged from standard 12-lead recordings to new single-lead data from smartphones and wearable patches. Clinical endpoints included the spectrum of cardiac dysfunction, including HFpEF, LVDD, and invasively confirmed elevated filling pressures.

Table 1. Summary of included studies.

Author	Year	TP	FN	FP	TN	AI/ML model type	ECG type	Primary endpoint
Kwon et al. [18]	2021	1,320	388	1,844	8,403	Deep learning	12-lead, 6-lead, single-lead	HFpEF
Sengupta et al. [19]	2018	107	26	9	46	Classical ML	12-lead signal-processed	Abnormal myocardial relaxation
Kuznetsova et al. [20]	2022	30	1	5	214	Classical ML	Single-lead (smartphone)	LVDD
Kagiyama et al. [21]	2020	51	25	62	250	Classical ML	12-lead signal-processed	LVDD/Abnormal relaxation
Lee et al. [22]	2024	18,205	3,678	13,136	63,744	Deep learning	12-lead	Increased filling pressure

Table 1. Summary of included studies. (continued)

Author	Year	TP	FN	FP	TN	AI/ML model type	ECG type	Primary endpoint
Sabovčik et al. [23]	2021	171	81	144	1,011	Classical ML	12-lead	LVDD
Unterhuber et al. [24]	2021	93	1	44	65	Deep learning	12-lead	HFpEF
Schlesinger et al. [25]	2025	836	393	848	2,543	Deep learning	Single-lead (wearable)	Elevated mPCWP
Gao et al. [26]	2025	43	17	16	41	Deep learning	12-lead	HFpEF risk (elevated LVEDP)

ECG: electrocardiogram; HFpEF: heart failure with preserved ejection fraction; LVDD: left ventricular diastolic dysfunction; ML: machine learning; AI: artificial intelligence.

Risk of bias assessment

The QUADAS-2 evaluation is summarized in Figure 2. All nine included studies were judged to have a low risk of bias in the *reference standard* domain, as each employed appropriate diagnostic comparators such as guideline-based echocardiography or invasive hemodynamic assessment. In contrast, *patient selection* was the most frequent source of concern. Several studies relied on retrospective designs, thus limiting the representativeness of their findings. Additional concerns were noted in the *flow and timing* domain when subsets of participants were excluded from final analyses, as a result, introducing selection bias.

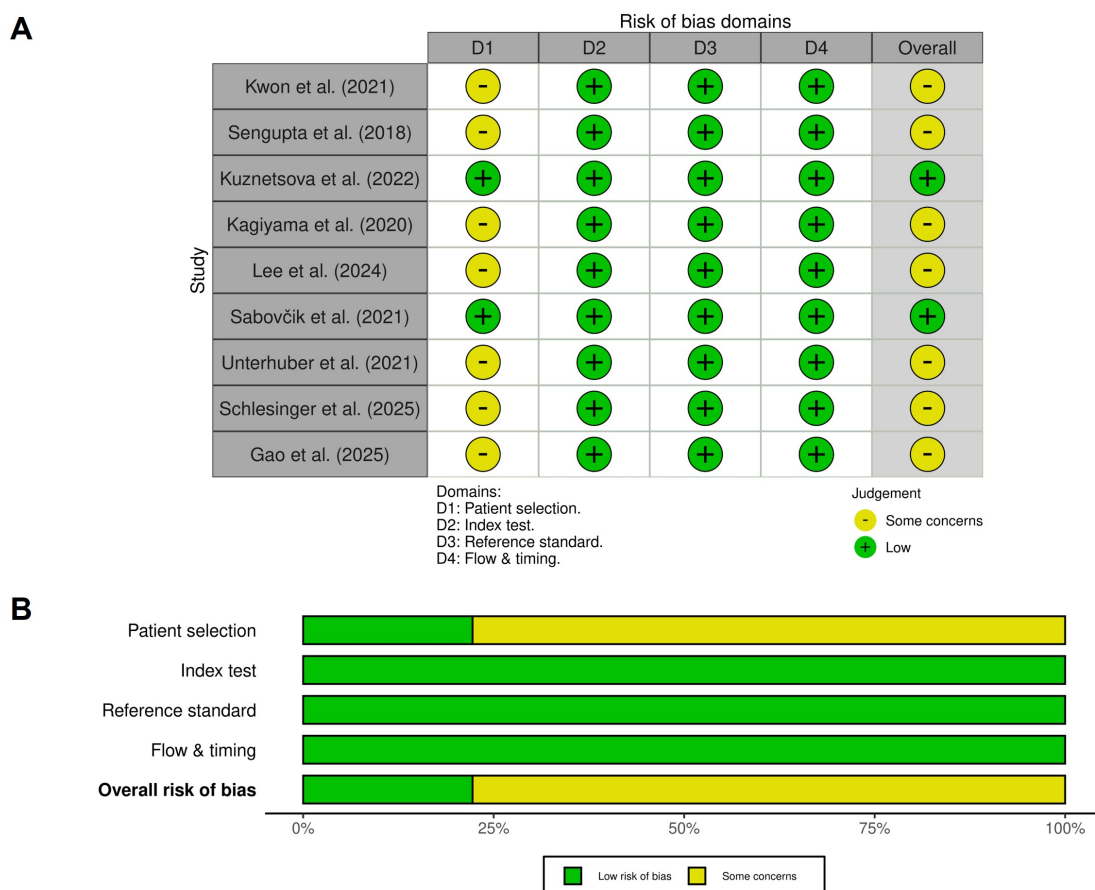


Figure 2. Risk of bias summary based on the QUADAS-2 assessment. Each row represents an included study, and columns represent the four domains of the QUADAS-2 tool (D1: patient selection, D2: index test, D3: reference standard, D4: flow & timing). Green indicates a low risk of bias, while yellow indicates some concerns.

Diagnostic accuracy

Forest plots of individual study estimates for sensitivity and specificity are shown in Figures 3 and 4. Reported sensitivities varied widely, ranging from 0.67 to 0.99 while specificities ranged from 0.60 to 0.98.

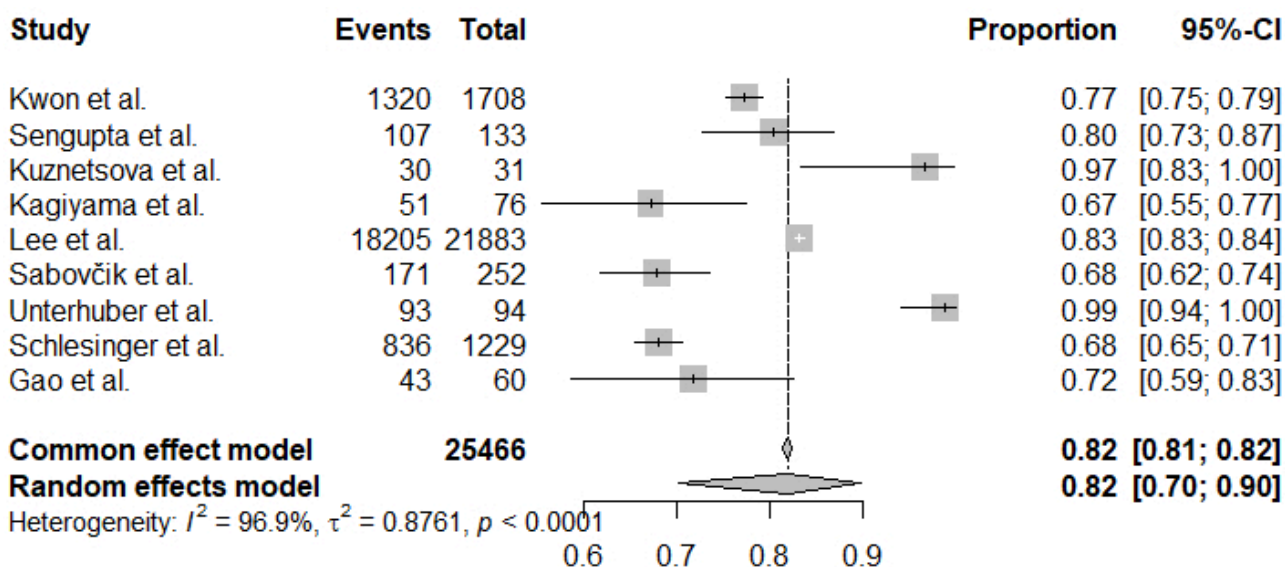


Figure 3. Forest plot of sensitivity for the detection of cardiac dysfunction. Each square represents the point estimate of sensitivity for an individual study, with the horizontal lines indicating the 95% confidence interval. A pooled estimate from the bivariate random-effects model is shown at the bottom.

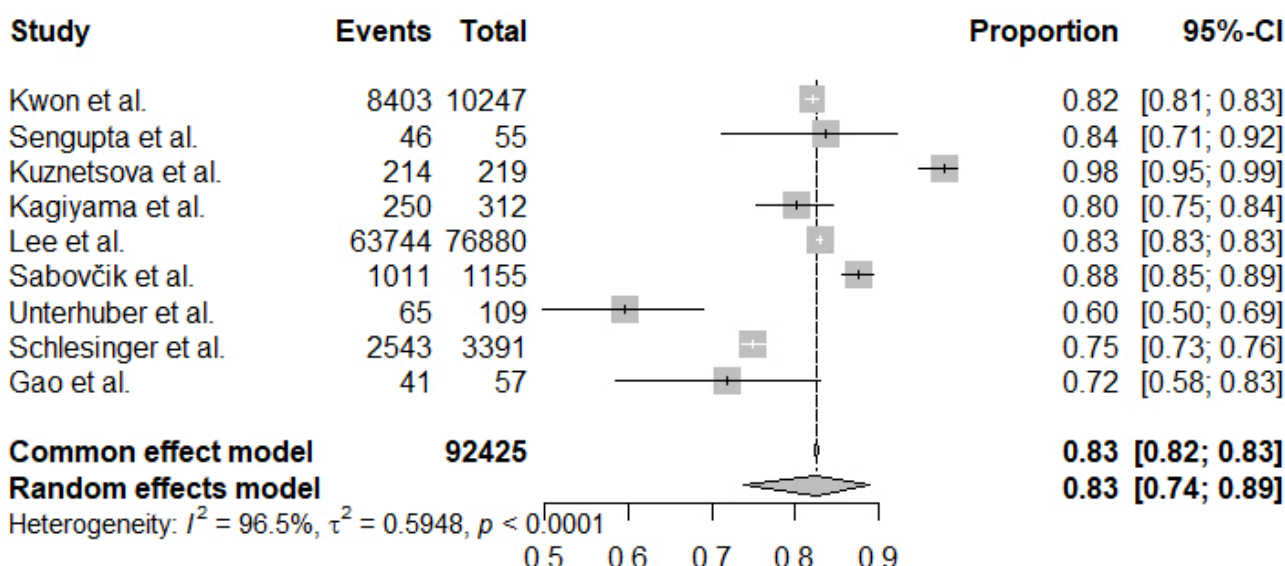


Figure 4. Forest plot of specificity for the detection of cardiac dysfunction. Each square represents the point estimate of specificity, proportional to its weight in the analysis. The horizontal lines indicate the 95% confidence interval. The diamond represents the pooled specificity of 0.83 [0.74, 0.89] from the random-effects model, with significant heterogeneity ($I^2 = 96.5\%$).

The bivariate random-effects meta-analysis produced a pooled sensitivity of 0.82 (95% CI: 0.70–0.90) and a pooled specificity of 0.83 (95% CI: 0.74–0.89). Both outcomes demonstrated extremely high and statistically significant heterogeneity ($I^2 > 96\%$, $p < 0.0001$).

To synthesize these findings and account for the correlation between sensitivity and specificity, a bivariate random-effects meta-analysis was performed. The resulting SROC curve is presented in Figure 5. The summary operating point for the AI-ECG models yielded a pooled sensitivity of 0.82 (95% CI: 0.70–0.90) and a pooled specificity of 0.83 (95% CI: 0.74–0.89). The 95% confidence region around this summary point indicates the precision of the mean estimate, while the much wider 95% prediction region illustrates the substantial heterogeneity between studies and the likely range of performance for a future study.

Investigation of heterogeneity

To investigate the cause of this heterogeneity, we conducted pre-specified subgroup analyses (Figures 6 and 7). We examined study-level specificity estimates according to AI model type (classical ML vs. DL). While individual studies showed variation in specificity, statistical comparison did not show a significant

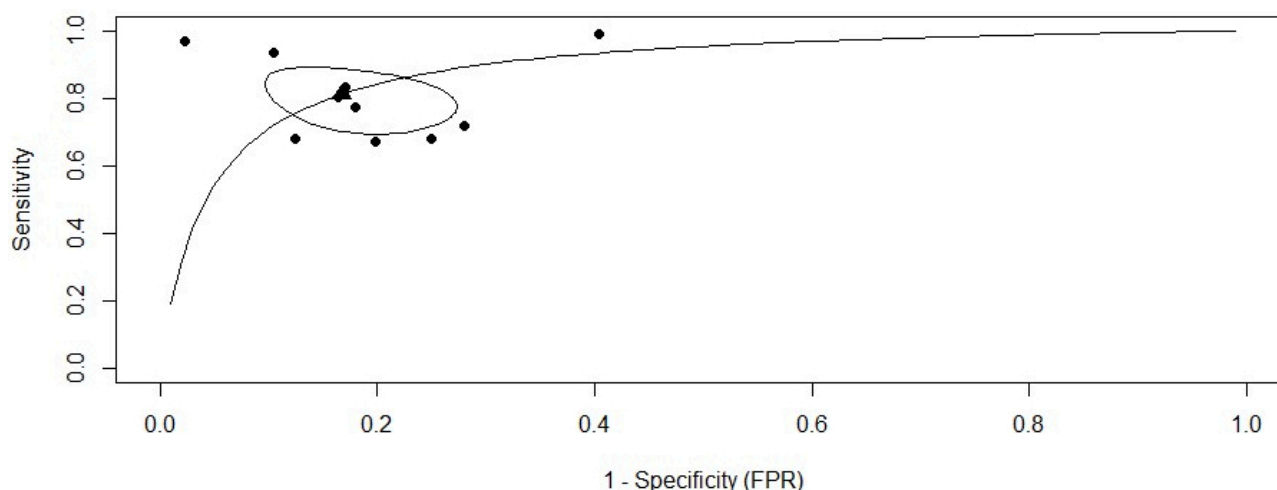


Figure 5. Summary receiver operating characteristic (SROC) curve for AI-ECG diagnostic accuracy. SROC curve illustrating the diagnostic performance of the included studies. Each black dot represents an individual study plotted according to its sensitivity and false positive rate (1-Specificity). The solid curve denotes the fitted SROC curve summarizing overall diagnostic accuracy. The black triangle indicates the pooled summary estimate of sensitivity and specificity, while the surrounding ellipse represents the 95% confidence region around this summary point. ECG: electrocardiogram; AI: artificial intelligence.

difference between model types ($p = 0.60$). High heterogeneity persisted in both groups. Likewise, subgrouping by clinical endpoint—diastolic dysfunction, HFpEF, or increased filling pressure—did not reveal significant between-group differences in specificity ($p = 0.52$). Importantly, heterogeneity remained very high across all subgroup comparisons ($I^2 > 77\%$), indicating that neither model type nor target condition accounted for the variability. Subgroup analyses for sensitivity yielded similar findings.

Discussion

In this systematic review and meta-analysis, we evaluated the diagnostic accuracy of nine different AI-ECG approaches for detecting cardiac dysfunction. Overall, the pooled results indicate a strong specificity (0.83) and a moderately high sensitivity (0.82). This pattern suggests that AI-ECG holds a considerable role as a non-invasive tool for ruling out cardiac dysfunction, which could help streamline diagnostic pathways and save echocardiography for patients who are most likely to benefit.

Yet, this encouraging finding is offset by the extremely high degree of heterogeneity we observed across studies. Our subgroup analyses, by model type (DL vs. conventional ML) and by clinical endpoint (HFpEF, LVDD, or raised filling pressures), could not explain this variability. The absence of a clear driver shows that model performance cannot be reduced to architecture or target condition alone; rather, it is shaped by a complex interplay of methodological and clinical factors.

Our assessment using QUADAS-2 sheds some light on some of these influences. The most common source of bias was patient selection. Several studies relied on retrospective or case-control designs, which may artificially inflate diagnostic accuracy due to spectrum bias and limited representation of real-world populations. In such settings, AI models are often trained and tested on highly selected cohorts with clear disease definitions, potentially overestimating performance when applied to broader clinical populations. In addition, trade-offs that are made during model development, such as prioritizing sensitivity over specificity, also contribute to these diverse results. For example, the model by Unterhuber et al. [24] was tuned to maximize sensitivity (0.99), but this came at the expense of specificity (0.60), making it most suitable as a rule-out test. In contrast, the algorithm from Kuznetsova et al. [20] achieved near-perfect specificity (0.98), better suited as a confirmatory tool. The extremely high heterogeneity observed across studies likely reflects multiple interacting factors beyond model architecture or diagnostic endpoint. These include variability in ECG acquisition (single-lead vs. 12-lead), signal preprocessing techniques, AI training strategies, disease prevalence, threshold selection, and reference standard definitions. Additionally, differences in population characteristics such as age distribution, comorbidity burden, and clinical setting

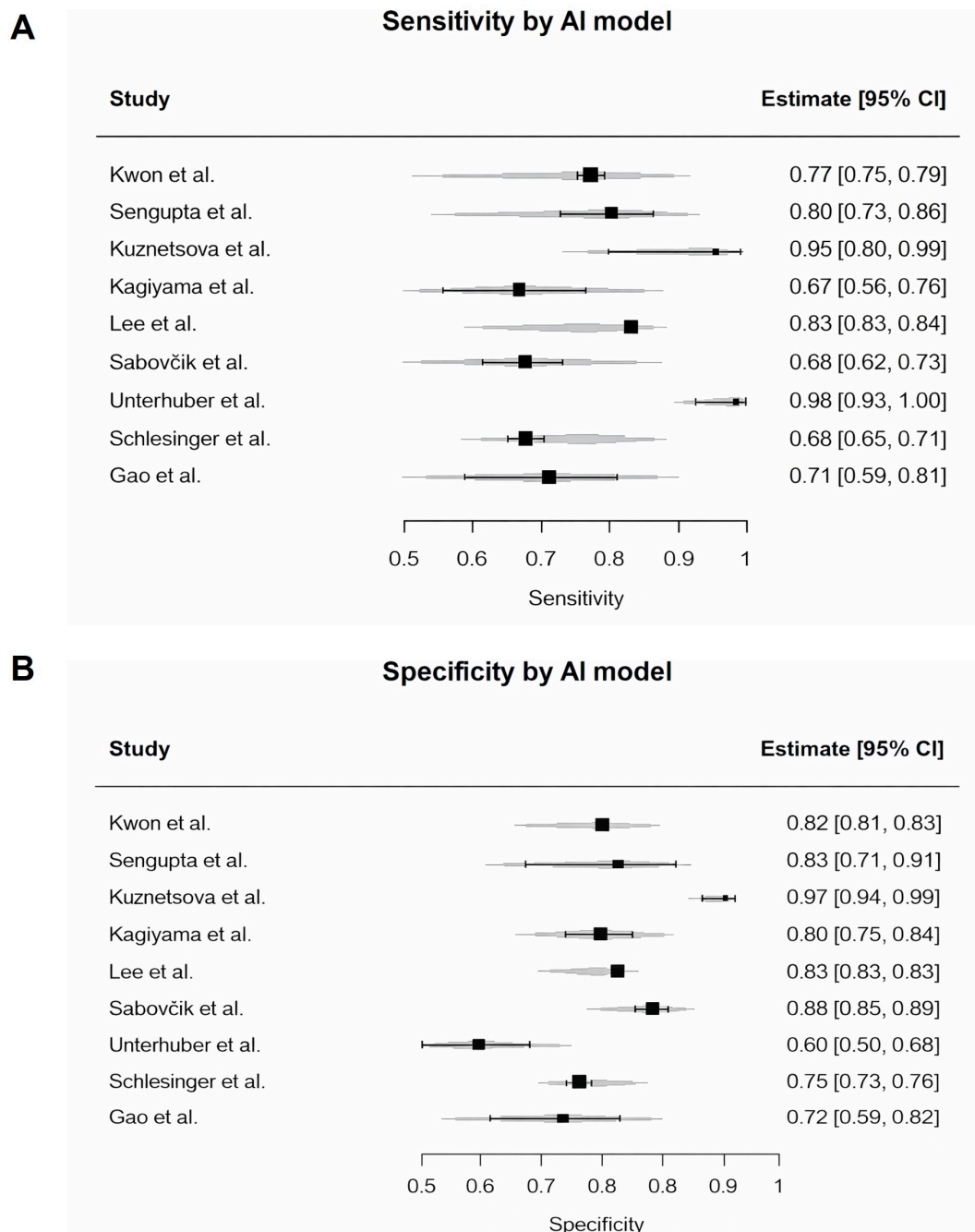


Figure 6. Study-level specificity estimates by AI model type. The figure shows individual study estimates for classical machine learning and deep learning models. Subgroup comparison was performed statistically and did not show a significant difference ($p = 0.60$). Heterogeneity remains high within both subgroups. AI: artificial intelligence.

may further contribute to variability. Such methodological and clinical diversity limits the interpretability of pooled estimates.

Clinical implications

Based on the available evidence, AI-enhanced ECG models appear most suitable as screening or triage tools rather than standalone diagnostic tests. Their high specificity and negative predictive value suggest potential utility in ruling out cardiac dysfunction in low- to intermediate-risk populations, thereby optimizing referral for echocardiography and improving resource allocation within clinical workflows.

Strengths and limitations

This meta-analysis has several strengths: a comprehensive synthesis of an emerging literature, a formal evaluation of study quality using QUADAS-2, and the application of a robust bivariate meta-analytic model.

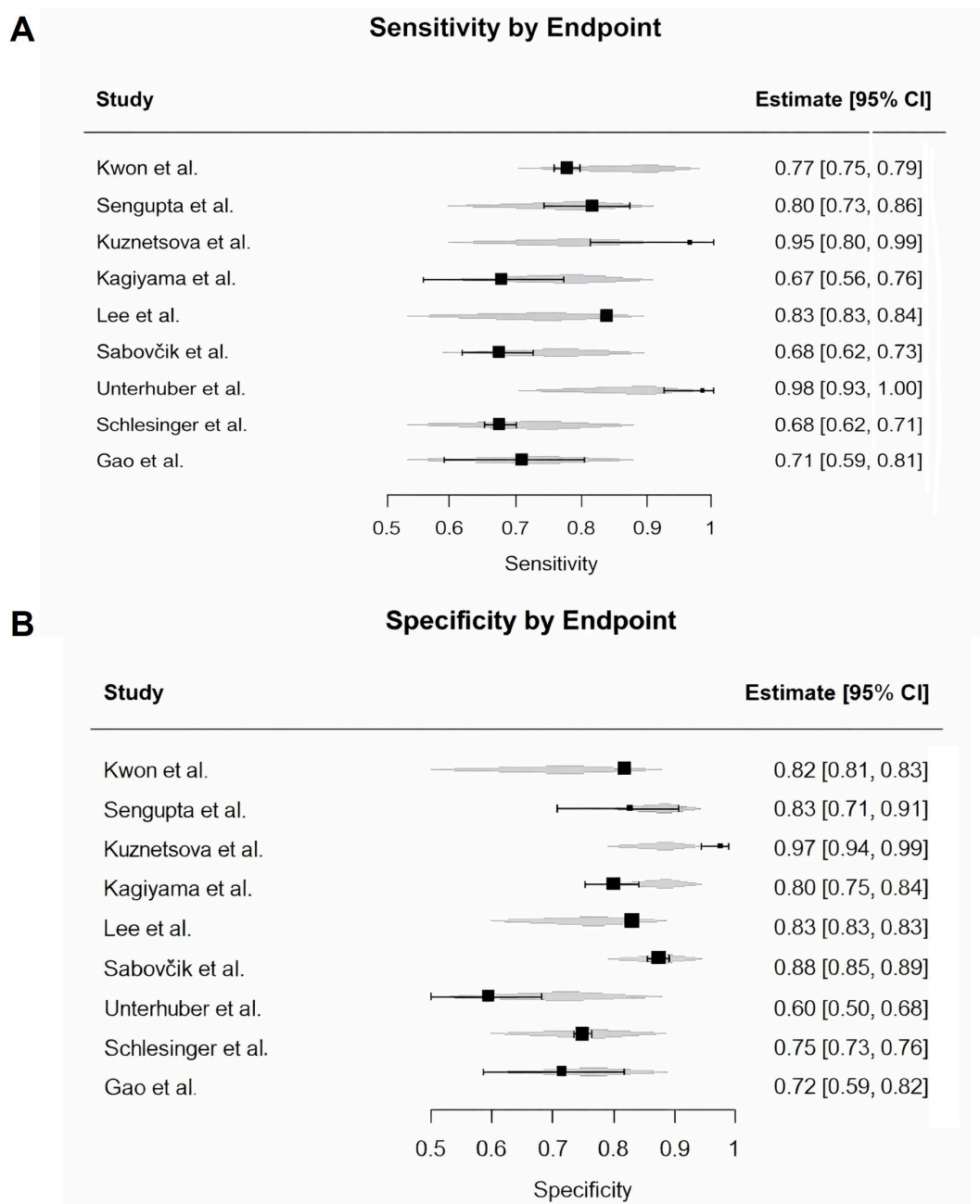


Figure 7. Subgroup analysis of specificity by clinical endpoint. The analysis revealed no statistically significant difference in specificity between studies targeting diastolic dysfunction, HFpEF, or increased filling pressure ($p = 0.52$). Substantial heterogeneity persists within each clinical category. HFpEF: heart failure with preserved ejection fraction.

Importantly, our exploration of heterogeneity, though inconclusive, highlights the methodological complexity underpinning this research field.

However, the findings must be interpreted with caution. The unexplained heterogeneity weakens the precision of the pooled estimates, which may represent little more than an average across very different contexts. The limited number of studies also reduced the power of subgroup comparisons. Also, the included studies were largely retrospective, and our dataset was based on a defined pool of studies rather than an exhaustive search, further limiting generalizability.

Conclusion

AI-enhanced ECG systems show high pooled specificity for detecting cardiac dysfunction and thus represent a promising, cost-effective, and accessible screening option. However, the field is marked by profound heterogeneity that cannot be explained by endpoint choice or model type alone, and model performance is

clearly context-dependent. Large-scale, prospective studies across diverse populations, along with better methodological standardization, are important for developing robust, generalizable AI-ECG tools that can be confidently integrated into clinical practice.

Abbreviations

AI: artificial intelligence

CI: confidence interval

DL: deep learning

ECG: electrocardiogram

HF: heart failure

HFpEF: heart failure with preserved ejection fraction

LVDD: left ventricular diastolic dysfunction

ML: machine learning

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SROC: summary receiver operating characteristic

Supplementary materials

The supplementary materials for this article are available at: https://www.explorationpub.com/uploads/Article/file/101293_sup_1.pdf.

Declarations

Author contributions

Habib S: Conceptualization, Investigation, Formal analysis, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing. Humna S: Conceptualization, Formal analysis, Writing—original draft, Writing—review & editing. AR: Conceptualization, Supervision, Methodology, Project administration, Writing—review & editing. HNT: Supervision, Project administration. Mursala T: Supervision, Project administration, Writing—review & editing. Mahad T: Conceptualization, Methodology, Writing—review & editing. Hooria S: Conceptualization, Methodology, Writing—review & editing. All authors read and approved the submitted version.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publication

Not applicable.

Availability of data and materials

The primary data for this meta-analysis were sourced online from databases listed in the methods. Referenced articles are accessible on PubMed, Cochrane, CTG and Google Scholar.

Funding

Not applicable.

Copyright

© The Author(s) 2026.

Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

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;118:3272–87. [DOI] [PubMed]
2. Escobar C, Palacios B, Varela L, Gutiérrez M, Duong M, Chen H, et al. Healthcare resource utilization and costs among patients with heart failure with preserved, mildly reduced, and reduced ejection fraction in Spain. *BMC Health Serv Res.* 2022;22:1241. [DOI] [PubMed] [PMC]
3. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al.; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. *Circulation.* 2020;141:e139–596. [DOI] [PubMed]
4. Escobar C, Palacios B, Varela L, Gutiérrez M, Duong M, Chen H, et al. Prevalence, Characteristics, Management and Outcomes of Patients with Heart Failure with Preserved, Mildly Reduced, and Reduced Ejection Fraction in Spain. *J Clin Med.* 2022;11:5199. [DOI] [PubMed] [PMC]
5. Iyngkaran P, Thomas MC, Neil C, Jelinek M, Cooper M, Horowitz JD, et al. The Heart Failure with Preserved Ejection Fraction Conundrum-Redefining the Problem and Finding Common Ground? *Curr Heart Fail Rep.* 2020;17:34–42. [DOI] [PubMed]
6. Budde H, Hassoun R, Mügge A, Kovács Á, Hamdani N. Current Understanding of Molecular Pathophysiology of Heart Failure With Preserved Ejection Fraction. *Front Physiol.* 2022;13:928232. [DOI] [PubMed] [PMC]
7. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, et al.; EMPEROR-Preserved Trial Investigators. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. *N Engl J Med.* 2021;385:1451–61. [DOI] [PubMed]
8. Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, et al.; DELIVER Trial Committees and Investigators. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. *N Engl J Med.* 2022;387:1089–98. [DOI] [PubMed]
9. Hagendorff A, Helfen A, Brandt R, Altiok E, Breithardt O, Haghi D, et al. Expert proposal to characterize cardiac diseases with normal or preserved left ventricular ejection fraction and symptoms of heart failure by comprehensive echocardiography. *Clin Res Cardiol.* 2023;112:1–38. [DOI] [PubMed] [PMC]
10. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al.; ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42:3599–726. [DOI] [PubMed]
11. Almeida DR, Andrade FA. Invasive Hemodynamic Monitoring in the Diagnosis of Heart Failure with Preserved Ejection Fraction. *ABC Heart Fail. Cardiomyop.* 2022;2:290–5. [DOI]
12. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25:30–6. [DOI] [PubMed] [PMC]

13. Harmon DM, Witt DR, Friedman PA, Attia ZI. Diagnosis and treatment of new heart failure with reduced ejection fraction by the artificial intelligence-enhanced electrocardiogram. *Cardiovasc Digit Health J*. 2021;2:282–4. [DOI] [PubMed] [PMC]
14. Zhao X, Zhang J, Gong Y, Xu L, Liu H, Wei S, et al. Reliable Detection of Myocardial Ischemia Using Machine Learning Based on Temporal-Spatial Characteristics of Electrocardiogram and Vectorcardiogram. *Front Physiol*. 2022;13:854191. [DOI] [PubMed] [PMC]
15. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. *Eur Heart J*. 2021;42:4717–30. [DOI] [PubMed] [PMC]
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. [DOI] [PubMed] [PMC]
17. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36. [DOI] [PubMed]
18. Kwon JM, Kim KH, Eisen HJ, Cho Y, Jeon KH, Lee SY, et al. Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features. *Eur Heart J Digit Health*. 2021;2:106–16. [DOI] [PubMed] [PMC]
19. Sengupta PP, Kulkarni H, Narula J. Prediction of Abnormal Myocardial Relaxation From Signal Processed Surface ECG. *J Am Coll Cardiol*. 2018;71:1650–60. [DOI] [PubMed]
20. Kuznetsova N, Gubina A, Sagirova Z, Dhif I, Gognieva D, Melnichuk A, et al. Left Ventricular Diastolic Dysfunction Screening by a Smartphone-Case Based on Single Lead ECG. *Clin Med Insights Cardiol*. 2022;16:11795468221120088. [DOI] [PubMed] [PMC]
21. Kagiya N, Piccirilli M, Yanamala N, Shrestha S, Farjo PD, Casaclang-Verzosa G, et al. Machine Learning Assessment of Left Ventricular Diastolic Function Based on Electrocardiographic Features. *J Am Coll Cardiol*. 2020;76:930–41. [DOI] [PubMed]
22. Lee E, Ito S, Miranda WR, Lopez-Jimenez F, Kane GC, Asirvatham SJ, et al. Artificial intelligence-enabled ECG for left ventricular diastolic function and filling pressure. *NPJ Digit Med*. 2024;7:4. [DOI] [PubMed] [PMC]
23. Sabovčik F, Cauwenberghs N, Kouznetsov D, Haddad F, Alonso-Betanzos A, Vens C, et al. Applying machine learning to detect early stages of cardiac remodelling and dysfunction. *Eur Heart J Cardiovasc Imaging*. 2021;22:1208–17. [DOI] [PubMed]
24. Unterhuber M, Rommel KP, Kresoja KP, Lurz J, Kornej J, Hindricks G, et al. Deep learning detects heart failure with preserved ejection fraction using a baseline electrocardiogram. *Eur Heart J Digit Health*. 2021;2:699–703. [DOI] [PubMed] [PMC]
25. Schlesinger DE, Alam R, Ringel R, Pomerantsev E, Devireddy S, Shah P, et al. Artificial intelligence for hemodynamic monitoring with a wearable electrocardiogram monitor. *Commun Med (Lond)*. 2025;5:4. [DOI] [PubMed] [PMC]
26. Gao Z, Yang Y, Yang Z, Zhang X, Liu C. Electrocardiograph analysis for risk assessment of heart failure with preserved ejection fraction: A deep learning model. *ESC Heart Fail*. 2025;12:631–9. [DOI] [PubMed] [PMC]