Original Article

Diagnostic Performance of Artificial Intelligence-Driven Electrocardiogram Algorithms in Detecting Heart Failure with Preserved Ejection Fraction: A Systematic Review and Meta-Analysis

Gustavo Barbosa David,[1] *Nicolas Ramos,*[1] *Ramon Huntermann,*[1] *Juan Peres de Oliveira,*[1] *Maria Eduarda Molinari,*[1] *Caroline de Oliveira Fischer Bacca*[1]

Centro Universitário para o Desenvolvimento do Alto Vale do Itajaí,[1] *Rio do Sul, SC – Brazil*

Abstract

Background: **Heart failure with preserved ejection fraction (HFpEF) is a highly prevalent syndrome associated with substantial morbidity and mortality. Accurate diagnosis is often challenging and costly. Recently, artificial intelligence-based electrocardiogram (AI-ECG) algorithms have emerged to make the diagnosis of HFpEF less expensive and more accessible. However, there is a lack of data regarding their accuracy.**

Objective: **We aimed to conduct a systematic review and meta-analysis to evaluate the diagnostic performance of AI-ECG algorithms in detecting HFpEF.**

Methods: **We searched PubMed, Embase, and Cochrane databases for studies evaluating AI-ECG algorithms in diagnosing HFpEF. We extracted true-positive, true-negative, false-positive, and false-negative events to estimate pooled sensitivity, specificity with 95% confidence intervals (CI), and area under the curve. Statistical analysis was performed using R under a random-effects model. Heterogeneity was assessed with I^2 statistics.**

Results: **We included three studies comprising 171,073 patients (66.86% with HFpEF), in which an AI-ECG was evaluated alongside echocardiography, with a left ventricular ejection fraction of ≥ 50%. The patients' mean age was 65.86 ± 4.5 years, and 48.12% were male. AI-ECG algorithms demonstrated a sensitivity of 0.89 (95% CI: 0.64 to 0.97), specificity of 0.80 (95% CI: 0.65 to 0.89), and an area under the receiver operating characteristic curve of 0.89 (95% CI: 0.85 to 0.92).**

Conclusion: **AI-ECG algorithms demonstrate moderate to high sensitivity, specificity, and accuracy in diagnosing HFpEF, supporting their potential for routine screening and early detection.**

Keywords: **Diastolic Heart Failure; Electrocardiography; Artificial Intelligence; Diagnosis.**

Introduction

Heart failure with preserved ejection fraction (HFpEF), characterized by a left ventricular ejection fraction (LVEF) ≥ 50%, affects approximately 3 million individuals in the United States and up to 32 million globally.[1]

Despite its high prevalence, HFpEF remains underdiagnosed, accounting for nearly half of all heart failure cases. It is associated with significant morbidity, frequent hospitalizations, and an annual mortality rate of around 15%. Early diagnosis and timely intervention are essential for optimizing management, improving patient outcomes, and reducing healthcare costs.[2,3]

Mailing Address: Nicolas Ramos •
Centro Universitário para o Desenvolvimento do Alto Vale do Itajaí. Dr. Guilherme Gemballa, 13. Postal code: 89160-932. Rio do Sul, SC – Brazil
E-mail: nicolas.ramos@unidavi.edu.br
Manuscript received May 15, 2025; revised manuscript October 29, 2025; accepted November 19, 2025.
Editor responsible for the review: Erito Marques

DOI: https://doi.org/10.36660/ijcs.20250091

Diagnosing HFpEF can be challenging due to its nonspecific symptoms and signs, often derived from patient history and physical examination. The electrocardiogram (ECG), however, is a valuable, noninvasive, and cost-effective tool in this context.[4] As pathophysiological changes in the heart influence its electrical activity, ECG data can provide crucial insights for early HFpEF detection.[5,6]

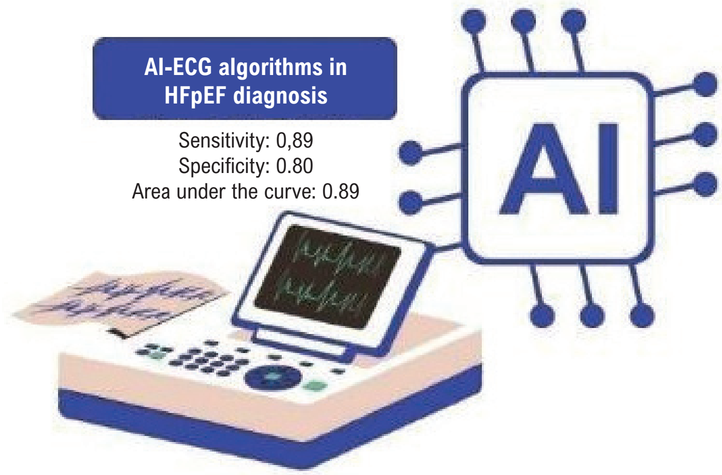
Recent advancements in artificial intelligence (AI), particularly machine learning and deep learning, have significantly improved diagnostic capabilities, enabling more accurate analysis of complex data.[7] When integrated with ECG, AI may have the potential to overcome diagnostic challenges in HFpEF.[8] This systematic review and meta-analysis aims to evaluate the diagnostic performance of artificial intelligence-based electrocardiogram (AI-ECG) algorithms in detecting HFpEF, providing a comprehensive assessment of their clinical utility.

Methods

We carried out this systematic review and meta-analysis following the guidelines outlined in the Cochrane

Central Illustration: **Diagnostic Performance of Artificial Intelligence-Driven Electrocardiogram Algorithms in Detecting Heart Failure with Preserved Ejection Fraction: A Systematic Review and Meta-Analysis**

INTERNATIONAL JOURNAL OF Cardiovascular SCIENCES



Int J Cardiovasc Sci. 2026; 39:e20250091

Source: Prepared by the authors. AI-ECG: artificial intelligence-based electrocardiogram; CI: confidence interval; HFpEF: heart failure with preserved ejection fraction.

Collaboration Handbook for Systematic Review of Interventions and the PRISMA Statement.[9,10] The protocol for the prospective meta-analysis was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the protocol number CRD420251002805.

Eligibility criteria

This meta-analysis was restricted to studies that met the following eligibility criteria: (1) observational studies, (2) evaluating patients who had both artificial-intelligence 12-lead ECG and echocardiograms conducted within a maximum interval of 30 days to assess LVEF, (3) population with HFpEF, and (4) reporting echocardiography as control method. There were no restrictions concerning the date or language of publication. Studies that included patients with heart failure with reduced ejection fraction (HFrEF), applied a cutoff different than 50% for HFpEF, or used magnetic resonance imaging as a control were excluded, as were abstracts and conference presentations.

Search and data extraction

We conducted a comprehensive search across PubMed, Embase, and Cochrane databases. Data were collected in January 2025, including all available studies up to that date. The following search terms were used in each of the three databases: ("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "neural networks") AND ("heart failure" OR "cardiac failure" OR CHF OR "congestive heart failure" OR HF OR HFpEF OR "heart failure with preserved ejection fraction" OR "diastolic heart failure") AND ("electrocardiogram" OR ECG OR EKG OR "electrocardiographic monitoring").

In addition, all references from the studies included and relevant reviews were manually examined to find more studies. Two authors conducted an independent analysis of the data using predefined search guidelines and quality evaluation methods. Disagreements between the two authors were resolved by consensus, including a third author.

Quality assessment

Risk of bias in the primary diagnostic studies included in our systematic review and meta-analysis was assessed independently by two review authors, using QUADAS-2.[11] All detailed questions regarding patient selection, the index test, the reference standard, as well as flow and timing were carefully reviewed and fully addressed. The evaluation of both the risk of bias and concerns regarding applicability in our review was categorized as high, unclear, or low. Any disagreements were resolved through discussion with a third review author.

Statistical analysis

We collected data on true-positive, true-negative, false-positive, and false-negative events as reported in the studies. These data were pooled to determine sensitivity, specificity with 95% confidence intervals (CI), and the area under the curve (AUC). Statistical analysis was conducted using R statistical software (version 4.2.2), employing a random-effects model.

Given the variability in baseline characteristics across studies, a leave-one-out sensitivity analysis was conducted to evaluate the impact of each individual study on the pooled estimates, ensuring that the overall findings were not disproportionately influenced by a single study.

Results

Our systematic review initially yielded 1,803 results. After removing duplicates and screening based on title and abstract, 20 full-text articles were reviewed for possible inclusion. Finally, three observational studies fulfilled our inclusion criteria and were included in the analysis, comprising a pooled population of 171,073 patients, of whom 114,383 (66.86%) were in the HFpEF group with a mean age of 65.86 ± 4.5 years. Details of the study selection are shown in Figure 1, and baseline characteristics are displayed in Table 1.

The meta-analysis revealed that AI-ECG demonstrated an estimated diagnostic sensitivity of 0.89 (95% CI: 0.64 to 0.97; Figure 2a) and a pooled specificity of 0.80 (95% CI: 0.65 to 0.89; Figure 2b) for HFpEF. Considerable heterogeneity was observed among the included studies (sensitivity: $I^2 = 99.4\%$; specificity: $I^2 = 99.8\%$). The summary receiver operating characteristic (SROC) analysis yielded an AUC of 0.89 (Figure 3). The main findings are shown in the Central Illustration.

Sensitivity analyses were performed using a leave-one-out approach (Supplementary Material Figures S1 and S2). The results demonstrated that the pooled estimates for both sensitivity and specificity remained consistent when each study was sequentially omitted, indicating that no single study had a disproportionate impact on the overall findings.

Quality assessment

Overall, the risk of bias was considered low in two studies (Vaid[12] and Kwon[13]) and unclear in one study (Unterhuber[14]). The domain-specific analyses conducted using the QUADAS-2 tool are summarized and presented in Table 2.

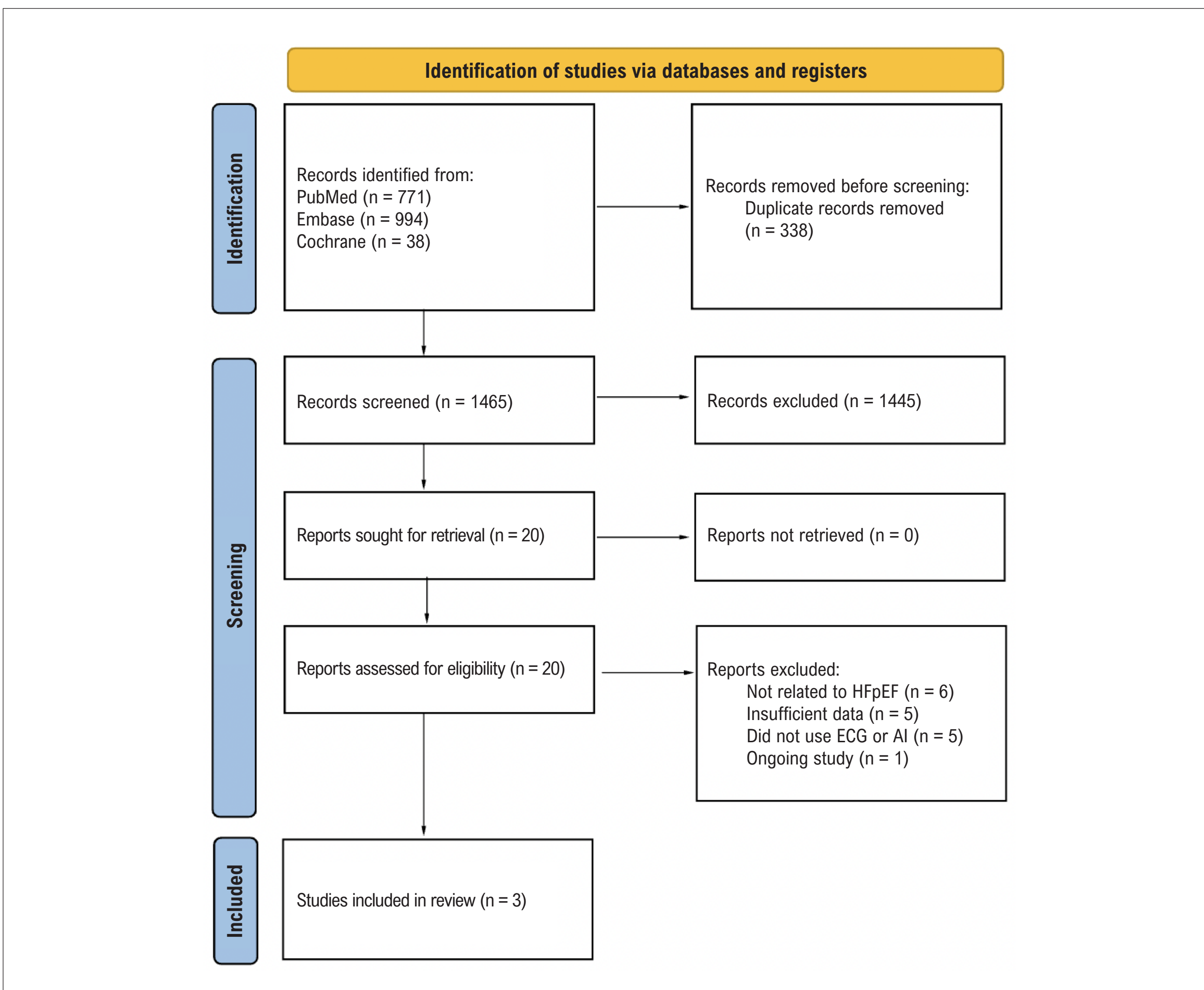


Figure 1 - *PRISMA flow diagram of study screening and selection. Source: Prepared by the authors. AI: artificial intelligence; ECG: electrocardiogram; HFpEF: heart failure with preserved ejection fraction.*

Table 1 – Baseline characteristics of included studies

Study	Type of AI	Standard diagnosis	Number of patients		Male sex, n (%)		Age, years (mean ± SD)		LVEF	
			HFpEF	Control	HFpEF	Control	HFpEF	Control	HFpEF	Control
Unterhuber 2021	Convolutional neural network	Echocardiogram; coronary angiography	720	1,164	390 (54.16)	746 (64.09)	66 ± 10	59 ± 10	NA	NA
Kwon 2020	Deep learning	Echocardiogram	4,048	18,100	56.70 (14.99)	70.44 (11.29)	56.70 ± 14.99	70.44 ± 11.29	57.14 ± 6.43	59.00 ± 6.26
Vaid 2022	Deep learning	Echocardiogram	109,615	37,426	55,831 (50.93)	25,358 (67.75)	65.63 ± 0.07	66.78 ± 5.36	62.0 ± 0	38.18 ± 3.49

Source: Prepared by the authors. AI: artificial intelligence; HFpEF: heart failure with preserved ejection fraction; LVEF: left ventricular ejection fraction; SD: standard deviation.

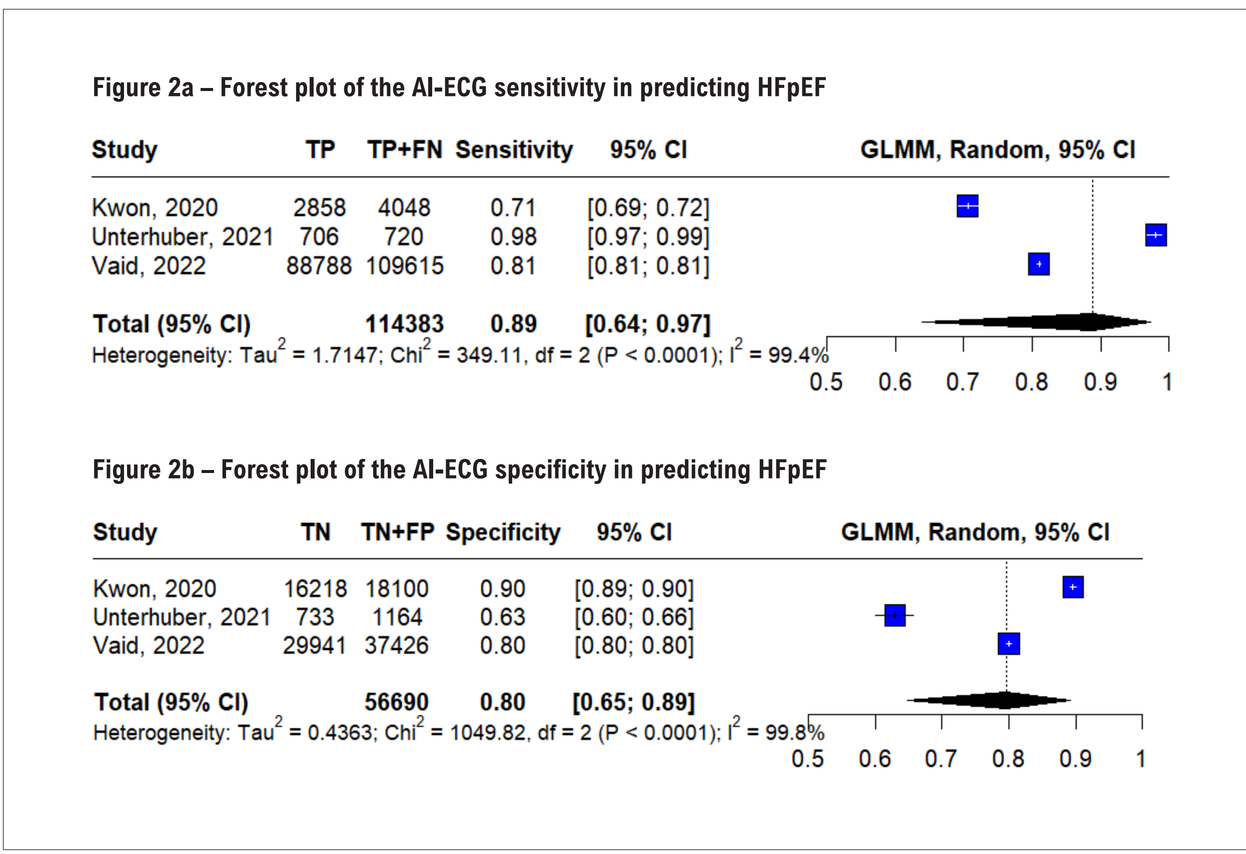


Figure 2 – *Source: Prepared by the authors. TP: True positive; FN: False negative; CI: Confidence interval; GLMM: Generalized linear mixed model*

Discussion

In this systematic review and meta-analysis of three studies, comprising 171,073 patients (66.86% with HFpEF), we assessed the diagnostic performance of AI-ECG models for detecting HFpEF, using echocardiography as the reference standard. The pooled data yielded the following main findings: (1) AI-ECG exhibited moderate to high sensitivity (0.89, 95% CI: 0.64 to 0.97) and specificity (0.80, 95% CI: 0.65 to 0.89); and (2) an overall AUC of 0.89 (95% CI: 0.85 to 0.92). These results show a strong discriminating capacity and a good balance between sensitivity and specificity. Moreover, the AUC indicates moderate to high accuracy for diagnosis according to the model, thus providing options for early detection and risk stratification regarding HFpEF.

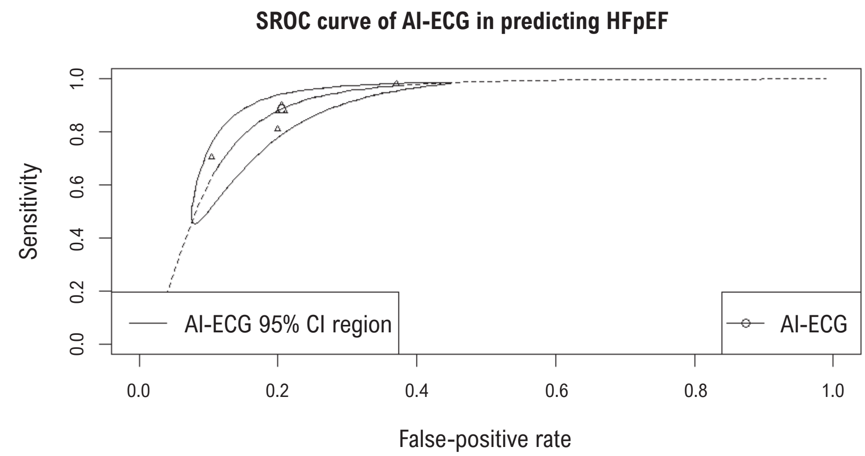


Figure 3 – *SROC curve of AI-ECG in predicting HFpEF. Source: Prepared by the authors. AI-ECG: artificial intelligence-based electrocardiogram; CI: confidence interval; HFpEF: heart failure with preserved ejection fraction.*

Table 2 – Risk of bias and applicability judgments in QUADAS-2

Study	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Unterhuber 2021	☹	☺	☺	?	☹	☺	☺
Kwon 2020	☹	☺	☺	☺	☹	☺	☺
Vaid 2022	☹	☺	☺	☺	☺	☺	☺

Source: Prepared by the authors. Notes: ☺: low risk; ☹: high risk; ?: unclear risk.

Our results coincide with other recent studies showing that deep learning models can efficiently detect minor changes in ECG waveform, enabling earlier identification of HFpEF as compared to conventional clinical ECG interpretation.[13] Notably, another study[14] reported a sensitivity of up to 99% when using a convolutional neural network applied to standard 12-lead ECGs in patients at risk for diastolic dysfunction, emphasizing the impressive rule-out capabilities of AI-ECG. These high-sensitivity models, however, may cause a higher false-positive rate, which emphasizes the need for confirmatory echocardiogram in ordinary clinical use.

Beyond basic diagnostic accuracy, AI-based ECG techniques offer promising prognostic potential. A recent study[15] revealed that patients with "false-positive" AI-ECG results were more likely to develop overt LV dysfunction on later follow-up, implying that even negative echocardiograms should require more attention if detected by AI. A subsequent study showed that combining AI-ECG outputs with basic clinical data might outperform some standard risk algorithms for asymptomatic LV dysfunction,[16] hence validating the theory that advanced computational tools detect subclinical alterations missed by typical ECG interpretations. Building on these observations, integrating clinical parameters—such as demographics, comorbidities, or laboratory biomarkers—with ECG-derived AI outputs may further improve diagnostic accuracy for HFpEF, resulting in more comprehensive and individualized prediction models. Growing data thus point to AI-ECG as a readily available triage tool, identifying people who might benefit from tailored therapy regimens and earlier echocardiographic investigation.

Our meta-analysis further shows that, for AI-ECG, the pooled estimates stayed consistently favorable even though different investigations used heterogeneous algorithms and datasets. This constancy emphasizes, even across different populations and healthcare settings, the strength of contemporary neural network designs for ECG interpretation.[17] Furthermore, subgroups of patients with questionable EF should particularly benefit from AI-driven screening since advanced models have shown the capacity to identify minor diastolic anomalies earlier than standard techniques.[18] These results highlight the possible synergy between AI-ECG and conventional therapeutic processes, therefore supporting our study's contribution to the existing body of knowledge.

Ferreira et al.[19] recently conducted a meta-analysis on the diagnostic accuracy of AI-ECG specifically for HFrEF (LVEF < 40%) rather than HFpEF. Their study similarly revealed strong pooled sensitivity (around 0.84) and specificity (about 0.89), despite concentrating on a clear EF cutoff, therefore supporting the general theory that AI-ECG may consistently identify underlying systolic dysfunction. Although their results highlight the value of AI-ECG in early diagnosis and risk classification for HFrEF, our meta-analysis adds the evidence regarding HFpEF, a clinical entity sometimes missed in standard tests. These two meta-analyses taken together show that AI-ECG has promise as a flexible, reasonably affordable screening tool spanning the range of EF—whether moderately lowered or preserved—highlighting the general .

Even with these promising findings, this study has some limitations. We tried to reduce selection bias by only adding studies with echocardiographic confirmation. Yet, residual confounding could still be introduced by the natural variation in ECG acquisition techniques and AI architectures. Furthermore, although we limited our meta-analysis to three trials satisfying tight inclusion criteria, the somewhat limited body of published evidence might constrain generalizability. Finally, none of the given datasets offered detailed cost-effectiveness or prospective outcome data; therefore, more studies are required to define actual therapeutic effects and clinical impact. Despite these constraints, our pooled results remain consistent with an increasing body of evidence, supporting AI-ECG as a reliable, non-invasive modality for HFpEF screening and early detection.

Conclusion

This meta-analysis demonstrates that AI-ECG algorithms exhibit moderate to high sensitivity, specificity, and diagnostic accuracy in detecting HFpEF, highlighting their potential as a valuable tool for early screening and diagnosis. These findings support the integration of AI-ECG algorithms into clinical practice as a promising approach for HFpEF detection, especially when combined with relevant clinical data to further enhance diagnostic accuracy. However, further large-scale prospective studies are necessary to validate these results and establish the optimal integration of AI-ECG technology across diverse healthcare settings .

Author Contributions

Conception and design of the research: David GB, Ramos N, Huntermann R, Bacca COF; acquisition of data: David GB, Ramos N; analysis and interpretation of the data: David GB, Ramos N, Huntermann R, Oliveira JP; statistical analysis: David GB, Huntermann R; writing of the manuscript: David GB, Ramos N, Huntermann R, Oliveira JP, Molinari ME; critical revision of the manuscript for intellectual content: David GB, Ramos N, Huntermann R, Oliveira JP, Bacca COF.

Potential Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Sources of Funding

There were no external funding sources for this study.

Study Association

This study is not associated with any thesis or dissertation work.

Ethics Approval and Consent to Participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Use of Artificial Intelligence

The authors did not use any artificial intelligence tools in the development of this work.

Availability of Research Data

The underlying content of the research text is contained within the manuscript.

References

1. Redfield MM, Borlaug BA. Heart Failure with Preserved Ejection Fraction: A Review. JAMA. 2023;329(10):827-38. doi: 10.1001/jama.2023.2020.
2. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in Prevalence and Outcome of Heart Failure with Preserved Ejection Fraction. N Engl J Med. 2006;355(3):251-9. doi: 10.1056/NEJMoa052256.
3. Wu J, Biswas D, Ryan M, Bernstein BS, Rizvi M, Fairhurst N, et al. Artificial Intelligence Methods foar Improved Detection of Undiagnosed Heart Failure with Preserved Ejection Fraction. Eur J Heart Fail. 2024;26(2):302-10. doi: 10.1002/ejhf.3115.
4. 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(5):282-4. doi: 10.1016/j.cvdhj.2021.08.002.
5. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, et al. Development and Validation of Deep-Learning Algorithm for Electrocardiography-Based Heart Failure Identification. Korean Circ J. 2019;49(7):629-39. doi: 10.4070/kcj.2018.0446.
6. Kwon JM, Jo YY, Lee SY, Kang S, Lim SY, Lee MS, et al. Artificial Intelligence-Enhanced Smartwatch ECG for Heart Failure-Reduced Ejection Fraction Detection by Generating 12-Lead ECG. Diagnostics. 2022;12(3):654. doi: 10.3390/diagnostics12030654.
7. 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(1):30-6. doi: 10.1038/s41591-018-0307-0.
8. 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: 10.3389/fphys.2022.854191.
9. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (Updated February 2022) [Internet]. London: Cochrane; 2022 [cited 2025 Feb 25]. Available from www.training.cochrane.org/handbook.

10. 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: 10.1136/bmj.n71.

11. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009.

12. Vaid A, Johnson KW, Badgeley MA, Somani SS, Bicak M, Landi I, et al. Using Deep-Learning Algorithms to Simultaneously Identify Right and Left Ventricular Dysfunction from the Electrocardiogram. JACC Cardiovasc Imaging. 2022;15(3):395-410. doi: 10.1016/j.jcmg.2021.08.004.

13. 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. 2020;2(1):106-16. doi: 10.1093/ehjdh/ztaa015.

14. 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(4):699-703. doi: 10.1093/ehjdh/ztab081.

15. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, et al. Prospective Validation of a Deep Learning Electrocardiogram Algorithm for the Detection of Left Ventricular Systolic Dysfunction. J Cardiovasc Electrophysiol. 2019;30(5):668-74. doi: 10.1111/jce.13889.

16. Potter EL, Rodrigues CHM, Ascher DB, Abhayaratna WP, Sengupta PP, Marwick TH. Machine Learning of ECG Waveforms to Improve Selection for Testing for Asymptomatic Left Ventricular Dysfunction. JACC Cardiovasc Imaging. 2021;14(10):1904-15. doi: 10.1016/j.jcmg.2021.04.020.

17. Dhingra LS, Aminorroaya A, Pedroso AF, Khunte A, Sangha V, McIntyre D, et al. Artificial Intelligence-Enabled Prediction of Heart Failure Risk from Single-Lead Electrocardiograms. JAMA Cardiol. 2025;10(6):574-84. doi: 10.1001/jamacardio.2025.0492.

18. Jentzer JC, Lee E, Attia Z, Hillerson D, Kane GC, Lopez-Jimenez F, et al. Artificial Intelligence ECG Diastolic Dysfunction and Survival in Cardiac Intensive Care Unit Patients. J Am Heart Assoc. 2025;14(5):e037839. doi: 10.1161/JAHA.124.037839.

19. Ferreira ALC, Feitoza LPGC, Benitez ME, Aziri B, Begic E, Souza LVF, et al. Diagnostic Accuracy of Artificial-Intelligence-Based Electrocardiogram Algorithm to Estimate Heart Failure with Reduced Ejection Fraction: A Systematic Review and Meta-Analysis. Curr Probl Cardiol. 2025;50(4):103004. doi: 10.1016/j.cpcardiol.2025.103004.