Mapping Algorithmic Bias in AI-Powered ECG Interpretation Across the AI Lifecycle: A Scoping Review Protocol

Luqman Lawal, Chris Paton, Mike English, Bruno Holthof, Tabitha Preston

Submitted to: JMIR Research Protocols
on: August 15, 2025

Table of Contents

Mapping Algorithmic Bias in AI-Powered ECG Interpretation Across the AI Lifecycle: A Scoping Review Protocol

Luqman Lawal[1]; Chris Paton[2] PhD, MBA; Mike English[1] MBChB; Bruno Holthof[1] MD, PhD, MBA; Tabitha Preston[1]

[1] Nuffield Departmet of Medicine Oxford University Oxford GB
[2] University of Oxford Oxford GB

Abstract

Background: Artificial intelligence (AI)–powered analysis of electrocardiograms (ECGs) is reshaping cardiac diagnostics, offering faster and often more accurate detection of conditions such as arrhythmias and heart failure. However, growing evidence suggests that algorithmic bias, defined as performance disparities across patient subgroups, may undermine diagnostic equity. These biases can emerge at any stage of the AI lifecycle, including data collection, model development, evaluation, deployment, and clinical use. If unaddressed, they risk exacerbating health disparities, particularly in underrepresented populations and low-resource settings. Early identification and mitigation of such bias are essential to ensuring diagnostic equity.

Objective: This scoping review aims to systematically map the current evidence on algorithmic bias in AI-enabled ECG interpretation. Specifically, we will (1) identify the types and sources of bias reported, (2) assess how performance varies across demographic or geographic subgroups, and (3) document any mitigation strategies applied. Our goal is to clarify how fairness is addressed in this growing field and highlight gaps that remain to be addressed to ensure equitable use of AI in cardiology.

Methods: We will conduct a comprehensive literature search across five electronic databases (PubMed, EMBASE, Cochrane CENTRAL, CINAHL, and IEEE Xplore), as well as grey literature sources including preprint servers and clinical trial registries. Eligible studies will include original research (2015–2025) evaluating the performance of AI-based ECG models across different subgroups or reporting on bias mitigation strategies. Two reviewers will independently screen studies, extract data using a standardized form, and resolve disagreements through consensus. The review will follow the PRISMA-ScR reporting framework.

Results: At the time of publication, study screening had not yet begun. Searches will commence in August 2025, with data extraction anticipated by September 2025. Results will be synthesized narratively and descriptively, and a PRISMA flow diagram will summarize the study selection process.

Conclusions: This review will be the first to comprehensively map the landscape of algorithmic bias in AI-powered ECG interpretation. By identifying patterns of inequity and evaluating proposed solutions, it will provide actionable insights for developers, clinicians, and policymakers aiming to promote fairness in AI-enabled cardiac care.

(JMIR Preprints 15/08/2025:82486)
DOI: https://doi.org/10.2196/preprints.82486

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

1.0 Title

Mapping Algorithmic Bias in AI-Powered ECG Interpretation Across the AI Lifecycle: A Scoping Review Protocol

2.0 Authors

Corresponding Author:
Dr. Luqman Lawal MD, MPH, MBA
Nuffield Department of Medicine, University of Oxford
Luqman.lawal@ndm.ox.ac.uk
+16513075288

Co-Authors and Affiliations:
Nuffield Department of Medicine, University of Oxford, UK
Dr Chris Paton
Prof. Mike English
Prof Bruno Holthof
Tabitha Preston

Table of Contents

3.0. Abstract

Background

Artificial intelligence (AI)–powered analysis of electrocardiograms (ECGs) is reshaping cardiac diagnostics, offering faster and often more accurate detection of conditions such as arrhythmias and heart failure. However, growing evidence suggests that algorithmic bias, defined as performance

disparities across patient subgroups, may undermine diagnostic equity. These biases can emerge at any stage of the AI lifecycle, including data collection, model development, evaluation, deployment, and clinical use. If unaddressed, they risk exacerbating health disparities, particularly in underrepresented populations and low-resource settings. Early identification and mitigation of such bias are essential to ensuring diagnostic equity.

Objective

This scoping review aims to systematically map the current evidence on algorithmic bias in AI-enabled ECG interpretation. Specifically, we will (1) identify the types and sources of bias reported, (2) assess how performance varies across demographic or geographic subgroups, and (3) document any mitigation strategies applied. Our goal is to clarify how fairness is addressed in this growing field and highlight gaps that remain to be addressed to ensure equitable use of AI in cardiology.

Methods

We will conduct a comprehensive literature search across five electronic databases (PubMed, EMBASE, Cochrane CENTRAL, CINAHL, and IEEE Xplore), as well as grey literature sources including preprint servers and clinical trial registries. Eligible studies will include original research (2015–2025) evaluating the performance of AI-based ECG models across different subgroups or reporting on bias mitigation strategies. Two reviewers will independently screen studies, extract data using a standardized form, and resolve disagreements through consensus. The review will follow the PRISMA-ScR reporting framework.

Results

At the time of publication, study screening had not yet begun. Searches will commence in August 2025, with data extraction anticipated by September 2025. Results will be synthesized narratively and descriptively, and a PRISMA flow diagram will summarize the study selection process.

Conclusions

This review will be the first to comprehensively map the landscape of algorithmic bias in AI-powered ECG interpretation. By identifying patterns of inequity and evaluating proposed solutions, it will provide actionable insights for developers, clinicians, and policymakers aiming to promote fairness in AI-enabled cardiac care.

Keywords

algorithmic bias; artificial intelligence; diagnostic fairness; electrocardiography; health equity; machine learning; cardiovascular diagnostics; scoping review

4.0 Introduction

4.1 Background

Artificial intelligence (AI)–enabled electrocardiogram (ECG) interpretation has emerged as a transformative tool in cardiovascular diagnostics. These models, particularly those built using deep learning architectures, have demonstrated high accuracy in identifying cardiac conditions such as atrial fibrillation, heart failure, and left ventricular dysfunction [1,2]. Given the global burden of cardiovascular disease (CVD), particularly in low-resource settings, AI-ECG tools offer potential to enhance access to timely diagnosis, especially in areas with limited cardiology expertise [3].

However, as AI adoption accelerates, concerns about algorithmic bias have gained prominence. Algorithmic bias refers to systematic disparities in model performance across subpopulations, often arising from unrepresentative training data, flawed modeling assumptions, or inappropriate deployment environments [4–6]. These biases can lead to unequal diagnostic accuracy and risk misdiagnosis, particularly in underrepresented or structurally marginalized groups. In cardiology, for instance, tools trained predominantly on White male datasets have underperformed in detecting myocardial infarction in women and individuals of African descent [7,8]. Physical diagnostic tools such as pulse oximeters have similarly demonstrated racial bias in measurement accuracy [9].

Bias may emerge across multiple stages of the AI development pipeline, from data collection and labeling to model training and evaluation, to deployment and clinical integration. Recent frameworks categorize these into five key stages: data, model, evaluation, deployment, and post-deployment bias [10]. Addressing each of these stages is critical to ensuring fairness, safety, and clinical utility. High-income countries (HICs) have started to institutionalize fairness evaluations in medical AI development, with regulatory and ethical guidelines now recommending transparency, community engagement, and

subgroup performance monitoring [11–13].

Despite this progress, significant gaps persist in understanding algorithmic bias within AI-ECG models, particularly in low- and middle-income countries (LMICs). Most existing studies and validations are based on datasets from North America and Europe, often excluding the very populations who might benefit most from AI-based screening tools[14]. This mismatch, described as “health data poverty” or “digital colonization,” raises concerns that AI tools exported to LMICs may underperform or cause unintended harm [15,16].

To date, there is no comprehensive synthesis of how bias manifests in AI-powered ECG interpretation. Existing reviews focus primarily on the diagnostic accuracy of AI-ECG for specific conditions and seldom report performance stratified by demographic or geographic subgroups [17]. A few studies have started to explore fairness or subgroup performance. Still, findings are scattered, and no prior scoping review has mapped the full range of bias-related evidence or mitigation efforts in this field. It remains unclear which populations are most affected, at what stages bias arises, and what strategies have been attempted to reduce disparities.

With the increasing deployment of AI-ECG tools, including in LMICs, there is an urgent need to map the evidence on algorithmic bias in this domain systematically. Doing so will support the equitable development and safe deployment of AI-based ECG technologies across diverse populations and health systems. Early identification and mitigation of such bias are essential to diagnostic equity.

4.2 Objectives

This scoping review aims to comprehensively map and characterize the current evidence related to algorithmic bias in AI-powered ECG interpretation across the AI lifecycle.

This review is guided by the **PCC (Population–Concept–Context)** framework, which also informs the eligibility criteria:
Population: Patients of all ages and demographic backgrounds undergoing ECG-based cardiac assessment, particularly those from underrepresented groups (e.g., women, racial/ethnic minorities, LMIC populations).
Concept: Algorithmic bias in AI-powered ECG interpretation, including disparities in diagnostic accuracy, error rates, and subgroup performance, as well as documented mitigation strategies.
Context: Any healthcare setting where AI-ECG models are developed, validated, or deployed, including both high-income and low-and middle-income countries.

By addressing the following research questions, this review aims to provide a comprehensive overview of the field:

1. What types of algorithmic bias are reported in AI-powered ECG interpretation models?

2. At which stages of the AI lifecycle (data, model, evaluation, deployment, post-deployment) do these biases occur?
3. How do identified biases affect diagnostic performance and healthcare equity across different patient groups?
4. What mitigation strategies or fairness interventions have been proposed or tested in the context of AI-ECG?
5. How does AI-ECG performance vary across demographic and geographic subgroups, including sex/gender, race/ethnicity, age, and LMIC vs. HIC settings?

This scoping review will inform future research, regulation, and implementation of AI-based cardiac diagnostics to ensure they serve all populations equitably.

5.0 Methods

5.1 Protocol Design

This study will be conducted as a scoping review, following the Arksey and O'Malley framework [18], with enhancements by Levac et al. [19] and methodological guidance from the Joanna Briggs Institute (JBI) [20]. The review will be reported following the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist [21].

We will undertake the five standard stages of scoping reviews:
1. Identifying the research question
2. Identifying relevant studies
3. Selecting studies
4. Charting the data
5. Collating, summarizing, and reporting results

We will skip the optional sixth stage (stakeholder consultation) because of resource and timing constraints.

5.2 Eligibility Criteria

Eligibility criteria are defined using the Population–Concept–Context (PCC) framework recommended for scoping reviews.

Population

We will include studies involving adult human participants (≥18 years) whose ECG data were analyzed using an AI-based tool. Studies must report subgroup performance data (e.g., by sex, race/ethnicity, age, or geographic setting), even if they do not explicitly mention "bias." Pediatric populations will be excluded.

Concept

The phenomenon of interest is algorithmic bias in AI-powered ECG interpretation. Eligible studies must evaluate AI models (e.g., machine learning, deep learning) used to interpret ECGs for cardiac diagnosis and must report on performance disparities, fairness metrics, or generalizability. Both commercial and research models will be included. Studies that assess model performance across groups, whether or not they use fairness terminology, will be eligible.

Context

We will include studies conducted in any global healthcare or research setting, including high-income and low- and middle-income countries (LMICs), across any level of care. No restrictions will be placed on the study setting or geographic region.

Study Designs

Eligible sources include all empirical studies presenting original data on AI-ECG performance and bias. These include:

- Diagnostic accuracy studies
- Randomized controlled trials (RCTs)
- Cohort and case–control studies
- Cross-sectional and retrospective validation studies
- Technical validation studies (e.g., internal or external validations of AI algorithms)

We will also include preprints, conference abstracts, and dissertations. Excluded sources include narrative/systematic reviews, commentaries, editorials, and purely theoretical/methodological papers without patient-level data.

Comparators

Studies with explicit or implicit comparators (e.g., comparing model outputs across subgroups or against clinicians) will be included. No comparator is required if subgroup-specific performance metrics are reported.

Time Frame

Studies published from January 1, 2015, to July 31, 2025, will be eligible. This window reflects the period during which deep learning models became prominent in ECG interpretation [22].

Language

Only studies published in English will be included. We acknowledge this as a limitation and will report any potentially relevant non-English studies excluded on this basis.

5.3 Information Sources and Search Strategy

A comprehensive search strategy will be developed in collaboration with a medical librarian and peer-reviewed using the PRESS checklist [23]. We will use both controlled vocabulary and free-text terms reflecting the PCC elements.

Databases

We will search the following databases:

- MEDLINE (via PubMed)
- EMBASE/ Ovid
- Cochrane CENTRAL
- CINAHL (EBSCO)
- IEEE Xplore
- Web of Science Core Collection (as supplemental sources)

Grey Literature

Additional sources include:

- ProQuest Dissertations & Theses
- Conference abstracts (e.g., AHA, ESC, AMIA)
- WHO ICTRP registries
- Reference lists of included studies and relevant reviews

Search Terms

Search strings will include terms such as:

(“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning”) AND (“ECG” OR “Electrocardiogram”) AND (“Bias” OR “Fairness” OR “Health Equity” OR “Disparities” OR “Subgroup Analysis”)

Search strategies will be customized for each database. Full strategies will be provided in an appendix. Filters will be applied for publication date (2015–2025) and English language.

5.4 Study Selection

Screening Procedure

Study selection will occur in two stages:

1. Title and abstract screening
2. Full-text screening

Two reviewers will screen independently. Discrepancies will be resolved by consensus or by a third reviewer. A pilot of 100 citations will calibrate inter-reviewer agreement.

Screening Tool

Covidence (Veritas Health Innovation) will be used to manage the screening process and document inclusion/exclusion decisions.

PRISMA Flow Diagram

Study selection will be reported using a PRISMA-ScR flowchart, including reasons for exclusion at the full-text stage.

5.5 Data Extraction (Charting the Data)

A structured data extraction form will be developed and piloted. Two reviewers will independently extract data from a sample of studies; one will complete the full extraction, with a second reviewer verifying all entries.

Data Items

Key variables to be extracted include:

- **Study characteristics**: Author, year, country, journal/source
- **Study design and setting**: Type, context (HIC vs LMIC, primary vs tertiary care)
- **Population**: Sample size, age, sex, race/ethnicity, comorbidities
- **AI model**: Algorithm type, clinical task, development dataset(s), model status (research vs commercial)
- **Performance and bias findings**: Stratified performance metrics (e.g., AUC, sensitivity/specificity), subgroup disparities, statistical significance
- **Bias lifecycle stage**: Data, model, evaluation, deployment, post-deployment (classified per Mehrabi et al. typology) [10]
- **Mitigation strategies**: Techniques such as resampling, fairness constraints, post hoc calibration, or external validation
- **Author's conclusions and implications**

Data will be stored in Excel or equivalent software for analysis.

5.6 Data Analysis and Synthesis

Synthesis Approach

We will use descriptive and thematic analysis. Results will be synthesized to address the five guiding questions of the review (Section 4.2). Quantitative summaries will describe study counts, subgroup types, and bias stages. Narrative synthesis will highlight:

- Bias types and lifecycle stages
- Performance disparities across subgroups
- Effectiveness of mitigation strategies

- Geographic and contextual representation (HIC vs LMIC)

Visualizations

Findings will be presented through:
- Evidence tables (e.g., bias type by study)
- Charts or maps (e.g., study distribution by year or geography)
- Summary matrices (e.g., subgroup performance data)

Gaps and Recommendations

We will identify evidence gaps (e.g., lack of LMIC data, underexplored bias stages) and recommend future research priorities. No meta-analysis will be performed due to expected heterogeneity.

5.7 Quality Appraisal

Consistent with JBI and PRISMA-ScR guidance, we will not conduct a formal quality appraisal. However, we will document evident methodological limitations (e.g., small sample sizes or limited subgroup analysis) that may affect interpretation.

6.0 Results:

Timeline
Database search: August 2025
Study screening and data extraction: August/September 2025
Analysis and manuscript submission: September 2025

Reporting Standards Compliance:
This protocol is reported following the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 guidelines.

Results (Present Status)

Progress to Date: August 2025
A preliminary search strategy has been developed in collaboration with an information specialist. Pilot searches were conducted to optimize search sensitivity and specificity. No formal screening, study selection, or data analysis has yet been performed.

Anticipated Completion:
We expect to finish the formal literature search and study selection process by August 2025. Data extraction, synthesis, and preparation of the full review manuscript are scheduled for completion by September 2025. These timelines are integrated into the broader work plan of the first author's D.Phil. research on AI-powered ECG interpretation in Sub-Saharan Africa.

Current Study Status: August 2025

This article describes a protocol. As of this writing, no study selection, inclusion assessment, data extraction, risk of bias assessment, or synthesis has been conducted. Therefore, the results section is not applicable at this stage.

No Actual Findings Yet

7.0. Discussion

This protocol outlines a robust and comprehensive approach to mapping the evidence on algorithmic bias in AI-powered ECG interpretation. By applying the PCC framework and PRISMA-ScR guidance, the study is designed to identify and categorize systematically reported biases, assess their impact on diagnostic performance, and review mitigation strategies across the AI lifecycle.

One of the main anticipated contributions of this review is its focus on both high-income and low- and middle-income country contexts, where the implications of algorithmic bias may differ substantially due to variations in data availability, clinical infrastructure, and deployment conditions. In doing so, the review addresses the persistent gap in global representativeness of AI-ECG research and the under-reporting of subgroup performance metrics.

This work is timely given the accelerating integration of AI in cardiovascular diagnostics and the increasing recognition that unchecked bias can exacerbate health inequities. Although previous reviews have examined the accuracy of AI-ECG systems, none have comprehensively mapped evidence on bias manifestations and mitigation efforts. By documenting both the problem space and potential solutions, our findings will provide actionable insights for developers, regulators, and clinicians.

We acknowledge certain limitations inherent to our protocol. Restricting inclusion to English-language publications may omit relevant studies, particularly from non-English-speaking LMICs. The expected heterogeneity in study designs and reporting standards will preclude meta-analysis, limiting the synthesis to descriptive and narrative methods. Additionally, the focus on published and indexed grey literature means that some proprietary or unpublished industry data, where bias assessments may have been conducted, will remain inaccessible.

Despite these constraints, our methodology incorporates multiple strategies to ensure comprehensiveness and transparency, including a librarian-led search strategy, independent screening by two reviewers, and standardized data extraction. The planned descriptive and thematic synthesis will enable a nuanced understanding of bias types, affected populations, and lifecycle stages, as well as the relative maturity of mitigation strategies.

Ultimately, the outputs of this review will serve as an evidence base for equitable AI-ECG development, deployment, and governance, helping to ensure that technological advancements in cardiology benefit all patient populations.

8.0 Ethics and Dissemination

Ethics

This review involves secondary analysis of publicly available literature and does not include any original research involving human participants. Therefore, it does not require approval from an institutional review board (IRB) or ethics committee. No individual-level, sensitive, or identifiable data will be accessed or analyzed during this study, consistent with established ethical standards for systematic reviews [23].

Dissemination

The results of this scoping review will be submitted for publication in a peer-reviewed journal focused on digital health, artificial intelligence in medicine, or cardiovascular care. In addition to traditional academic dissemination through journal articles and conference presentations, findings will also be shared with key stakeholders, such as healthcare AI developers, clinicians, and policymakers, via webinars, social media, and professional networks. Special emphasis will be placed on sharing the findings in ways that are accessible to stakeholders in low- and middle-income countries, where bias in AI tools can be particularly impactful.

9.0. Funding

This is part of a self-funded D.Phil research work conducted by the first author at the University of Oxford.

10.0 Conflicts of Interest

The authors declare no competing interests related to this scoping review protocol.

11.0 Acknowledgements

The authors gratefully acknowledge the support of the University of Oxford’s Nuffield Department of Primary Health Care Sciences and the guidance of the D.Phil. supervisory team. We also thank the Librarian we consulted during the development of the search strategy. This work is part of a doctoral research program supported by academic training and collaboration with the Mayo Clinic Platform and other global digital health partners.

12.0 References

1. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet. 2019;394(10201):861–867.

2. Kwon JM, Lee SY, Jeon KH, et al. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. Scand J Trauma Resusc Emerg Med. 2020;28(1):98.

3. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–453.

4. Pfohl SR, Foryciarz A, Shah NH. An empirical characterization of fair machine learning for clinical risk prediction. J Biomed Inform. 2021;113:103621.

5. Dey S, Luo Y, Barman P, et al. Bias in biosignals: A case study on pulse oximetry. NPJ Digit Med. 2022;5(1):10.

6. Wijns W, Kolh P, Danchin N, et al. Guidelines on myocardial revascularization: the Task Force on Myocardial Revascularization of the European Society of Cardiology (ESC). Eur Heart J. 2010;31(20):2501–2555.

7. Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics. 2019;21(2):E167–179.

8. Sjoding MW, Dickson RP, Iwashyna TJ, et al. Racial bias in pulse oximetry measurement. N Engl J Med. 2020;383(25):2477–2478.

9. Mehrabi N, Morstatter F, Saxena N, et al. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54(6):1–35.

10. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health. 2019;9(2):020318.

11. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep. 2019;49(1):15–21.

12. Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. N Engl J Med. 2017;376(26):2507–2509.

13. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device Action Plan. 2021.

14. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: WHO; 2021.

15. Arora A, Yao CA, Stiegel KR, et al. Reimagining AI for Africa: algorithmic bias, health data poverty and digital colonization. NPJ Digit Med. 2022;5(1):1–4.

16. Weissglass D, Owoyemi O, Mahomed S, et al. Ethical artificial intelligence in sub-Saharan Africa: challenges and opportunities. Health Policy Plan. 2022;37(7):831–839.

17. Mittelstadt B, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics. 2016;22(2):303–341.

18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19–32.

19. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5:69.

20. Peters MDJ, Godfrey C, McInerney P, et al. Chapter 11: Scoping Reviews. In: Aromataris E, Munn Z, eds. JBI Manual for Evidence Synthesis. Joanna Briggs Institute; 2020.

21. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169(7):467–473.

22. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25(1):65–69.

23. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol. 2016;75:40–46.

13.0 Appendix

1. Database Search Strings

2. Data Extraction Template

Variable	Definition / Coding Instructions
Study ID	Unique identifier assigned after screening.
Citation	Author(s), year, journal/conference, DOI.
Country	World Bank classification of country (HIC, UMIC, LMIC, LIC).
Setting	Clinical setting (e.g., tertiary hospital, outpatient clinic, community health center).
Income Level	World Bank classification of country (HIC, UMIC, LMIC, LIC).
Study Design	RCT, observational, case-control, diagnostic accuracy, validation, simulation, etc.
AI Model Type	Architecture (e.g., CNN, RNN, hybrid), proprietary vs. open-source.
Training Data Source	Dataset name, origin, size, and population characteristics.
Test Data Source	Dataset name, origin, size, and population characteristics.
Bias Type	Using Mehrabi et al.'s typology (sampling bias, measurement bias, etc.).
AI Lifecycle Stage	Stage(s) where bias was reported (data collection, preprocessing, modeling, deployment).
Performance Metrics Reported	AUROC, sensitivity, specificity, PPV, NPV, F1, etc.
Subgroup Analyses	Whether performance metrics were stratified by demographics or comorbidities.
Mitigation Strategies	Bias mitigation approaches described (e.g., data augmentation, reweighting).
Regulatory / Ethical Considerations	Any discussion of fairness, equity, or compliance with standards.
Key Findings	Summary of main results related to bias.
Limitations Reported	Limitations acknowledged by the authors regarding bias.
Reviewer Notes	Free-text space for reviewer comments.

Database	Search String
MEDLINE (PubMed)	(("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "Deep Learning"[Mesh] OR "artificial intelligence"[tiab] OR "machine learning"[tiab] OR "deep learning"[tiab] OR "neural network*"[tiab] OR algorithm*[tiab]) AND ("Electrocardiography"[Mesh] OR "Electrocardiography, Ambulatory"[Mesh] OR ECG[tiab] OR EKG[tiab] OR "electrocardiogram*"[tiab]) AND ("Health Equity"[Mesh] OR "Healthcare Disparities"[Mesh] OR "Bias"[Mesh] OR "Prejudice"[Mesh] OR fairness[tiab] OR "health equity"[tiab] OR disparit*[tiab] OR "subgroup analy*"[tiab] OR generalizability[tiab] OR "performance variation"[tiab])) AND ("2015/01/01"[Date - Publication] : "3000"[Date - Publication]) AND English[Language]
EMBASE (Ovid)	(('artificial intelligence'/exp OR 'machine learning'/exp OR 'deep learning'/exp OR 'artificial intelligence':ti,ab OR 'machine learning':ti,ab OR 'deep learning':ti,ab OR 'neural network*':ti,ab OR algorithm*:ti,ab) AND ('electrocardiography'/exp OR ECG:ti,ab OR EKG:ti,ab OR electrocardiogram*:ti,ab) AND ('health equity'/exp OR 'health care disparity'/exp OR 'bias'/exp OR 'prejudice'/exp OR fairness:ti,ab OR 'health equity':ti,ab OR disparit*:ti,ab OR 'subgroup analy*':ti,ab OR generalizability:ti,ab OR 'performance variation':ti,ab)) AND [english]/lim AND [2015-2025]/py
CINAHL (EBSCO)	((MH "Artificial Intelligence+") OR (MH "Machine Learning+") OR (MH "Deep Learning+") OR TI,AB("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR algorithm*)) AND ((MH "Electrocardiography+") OR TI,AB("ECG" OR "EKG" OR "electrocardiogram*")) AND ((MH "Health Equity+") OR (MH "Healthcare Disparities+") OR (MH "Bias+") OR (MH "Prejudice+") OR TI,AB(fairness OR "health equity" OR disparit* OR "subgroup analy*" OR generalizability OR "performance variation"))) AND (LA English) AND (DT 20150101-20251231)
IEEE Xplore	(("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR algorithm*) AND ("ECG" OR "EKG" OR "electrocardiogram*") AND (fairness OR "health equity" OR disparit* OR "subgroup analy*" OR generalizability OR "performance variation")) AND (Publication Year: 2015-2025)
Web of Science (Core Collection)	TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR algorithm*) AND TS=("ECG" OR "EKG" OR "electrocardiogram*") AND TS=(fairness OR "health equity" OR disparit* OR "subgroup analy*" OR generalizability OR "performance variation") Refined by: Languages=(ENGLISH) AND Publication Years=(2015-2025)

3. PRISMA -ScR

Section	Item No.	Checklist Item	Location in Protocol
TITLE	1	Identify the report as a scoping review.	1
ABSTRACT	2	Provide a structured summary including background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	3, 4
INTRODUCTION	3	Describe the rationale for the review in the context of what is already known.	5
	4	Provide an explicit statement of the objectives or questions the review addresses.	6

METHODS	5	Indicate whether a review protocol exists, if and where it can be accessed	7
	6	Specify characteristics of the sources of evidence (e.g., years considered, language, and publication status) and eligibility criteria used for the review.	8
	7	Describe all information sources (e.g., databases with dates of coverage, contact with authors) in the search and date last searched.	9
	8	Present the full electronic search strategy for at least one database, including any limits used, so it could be repeated.	9
	9	State the process for selecting sources of evidence (screening and eligibility).	9, 10
	10	Describe the methods of charting data from the included sources (data extraction) and any processes for obtaining or confirming data from investigators.	10
	11	List and define all variables for which data were sought and any assumptions or simplifications made.	10
	12	Describe the methods used for critical appraisal of individual sources of evidence (if done), including how this information is to be used in any data synthesis.	11
	13	Describe the methods of handling and summarizing the data that were charted.	11
RESULTS	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	12
	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	12
	16	Present the results of any critical appraisal of individual sources of evidence (if done).	12
	17	Present relevant data that were charted that relate to the review questions and objectives.	12
	18	Summarize and/or present the results of the synthesis.	12
DISCUSSION	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	12
	20	Discuss the limitations of the scoping review process.	13
	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	13
FUNDING	22	Describe sources of funding for the included sources of evidence and for the scoping review itself. Describe the role of the funders.	14

4. Bias Lifecycle Stage Framework

Bias Type (Mehrabi et al.)	Operational Definition in This Review	Lifecycle Stage
Sampling Bias	Systematic differences in patient populations included in the dataset vs. target population.	Data Collection
Measurement Bias	Errors in ECG acquisition or labeling that disproportionately affect subgroups.	Data Collection / Preprocessing
Label Bias	Inaccuracies or inconsistencies in ground truth (e.g., cardiologist interpretation differences).	Data Annotation
Aggregation Bias	Ignoring relevant subgroup differences during modeling, leading to one-size-fits-all models.	Modeling
Evaluation Bias	Incomplete or skewed validation datasets that do not represent target population diversity.	Model Evaluation
Deployment Bias	Performance degradation or inequity during real-world use due to context differences.	Deployment

5. Definition of Terminologies

Term	Definition
Algorithmic Bias	Systematic and unfair performance differences of an AI model between different subgroups of patients.
Fairness	The principle of achieving equitable diagnostic performance across patient subgroups.
AI Lifecycle Stage	Phases from data collection to deployment, as defined by model development workflows.
ECG Interpretation	The process of analyzing electrocardiogram data to detect cardiac conditions.