INPLASY

Diagnostic Performance of Artificial Intelligence Algorithms in Predicting Atrial Fibrillation in Patients with Sinus Rhythm on Electrocardiograms: A Systematic Review and Meta-Analysis

INPLASY202650151
doi: 10.37766/inplasy2026.5.0151
Received: 27 May 2026
Published: 27 May 2026

Corresponding author:
WANG Xinkang

13305027368@163.com

Author Affiliation:
Shengli Clinical Medical College, Fujian Medical University.

Huang, YX; Zhang, XL; Wang, XK.

ADMINISTRATIVE INFORMATION

Support - Joint Funds for the innovation of science and Technology, Fujian province(Grant number :2025Y9031).

Review Stage at time of this submission - Data extraction.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202650151

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 27 May 2026 and was last updated on 27 May 2026.

INTRODUCTION

Review question / Objective Population
Adult patients (≥18 years of age) with baseline 12-lead sinus rhythm electrocardiograms (SR-ECG) recorded prior to any clinical diagnosis of atrial fibrillation.
Intervention
Application of artificial intelligence (AI) algorithms, including machine learning or deep learning models, to surface standard ECG traces for the detection of underlying paroxysmal, persistent, or permanent atrial fibrillation.
Comparator
The comparator was the clinical reference standard, defined as documented AF during follow-up (via ECG, Holter monitoring, implantable loop recorder, or medical record adjudication). If studies additionally compared the discrimination of AI models against clinical risk scores, the CHARGE-AF score was extracted for head-to-head AUC comparison.
Outcome
The primary outcomes were diagnostic accuracy metrics, including sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). Studies reporting raw contingency data (true positives, false positives, false negatives, true negatives) sufficient to calculate these metrics were also included. Additionally, hazard ratios with 95% confidence intervals for incident atrial fibrillation based on AI-predicted risk were extracted.
Study Characteristics
Original peer-reviewed research articles in English with full-text availability, employing cohort, case-control diagnostic designs, or randomized controlled trials. Eligible studies were required to validate the AI model on an independent test dataset that was strictly held out from all stages of model training and hyperparameter tuning.
Other
Studies focusing solely on ECG signal analysis without integration of non-ECG clinical parameters (e.g., biomarkers, demographics) into the model input.

Condition being studied Atrial Fibrillation.

METHODS

Participant or population Adult patients (≥18 years of age) with baseline 12-lead sinus rhythm electrocardiograms (SR-ECG) recorded prior to any clinical diagnosis of atrial fibrillation.

Intervention Application of artificial intelligence (AI) algorithms, including machine learning or deep learning models, to surface standard ECG traces for the detection of underlying paroxysmal, persistent, or permanent atrial fibrillation.

Comparator The comparator was the clinical reference standard, defined as documented AF during follow-up (via ECG, Holter monitoring, implantable loop recorder, or medical record adjudication). If studies additionally compared the discrimination of AI models against clinical risk scores, the CHARGE-AF score was extracted for head-to-head AUC comparison.

Study designs to be included Original peer-reviewed research articles in English with full-text availability, employing cohort, case-control diagnostic designs, or randomized controlled trials. Eligible studies were required to validate the AI model on an independent test dataset that was strictly held out from all stages of model training and hyperparameter tuning.

Eligibility criteria Inclusion criteria include:
(1) Participants: Adult patients (≥18 years of age) with a baseline 12-lead sinus rhythm electrocardiogram (SR-ECG) recorded prior to any clinical diagnosis or documented history of atrial fibrillation.
(2) Interventions: The application of an artificial intelligence algorithm, including but not limited to machine learning or deep learning models, to analyze the raw or pre-processed digital signal of the standard, surface 12-lead ECG recorded during sinus rhythm. The intervention is the AI model's output (e.g., a prediction, probability score, or classification) indicating the risk or presence of underlying atrial fibrillation (paroxysmal, persistent, or permanent).
(3) Comparator: The clinical reference standard for the definitive diagnosis of atrial fibrillation. This must be documented AF occurring during a follow-up period after the baseline ECG, confirmed by one or more of the following methods: a subsequent standard 12-lead ECG demonstrating AF, continuous ambulatory monitoring, or adjudication via medical record review (e.g., physician diagnosis, hospital discharge codes, arrhythmia confirmed in clinical notes).
(4) Outcomes: Studies must report at least one of the following diagnostic accuracy metrics for the AI algorithm's performance in detecting underlying AF from the sinus rhythm ECG: sensitivity, specificity, or the area under the receiver operating characteristic curve (AUC). Alternatively, studies must provide the raw contingency table data (counts of true positives, false positives, false negatives, and true negatives) from an independent test set, which is sufficient for the meta-analysis to calculate these metrics.
(5) Study Design: Original, peer-reviewed research articles published in English with full-text availability. Eligible study designs include diagnostic cohort studies (prospective or retrospective), diagnostic case-control studies, or randomized controlled trials (RCTs) with a diagnostic accuracy component. A critical and mandatory requirement is that the study must validate the final AI model on a strictly independent test dataset. This test set must have been completely held out during all stages of model training, development, and hyperparameter tuning.
Exclusion criteria include:
(1) Studies where the full text of the article is not accessible or cannot be obtained through reasonable means.
(2) Studies that develop or validate AI models which integrate non-ECG clinical parameters (e.g., patient demographics, laboratory biomarkers, clinical risk scores) as part of the model's input features for making the AF prediction from the sinus rhythm ECG. This criterion excludes models that are not based solely on ECG signal analysis.
(3) Studies that do not provide sufficient detail to confirm that the independent test set was strictly held out from all model development processes (training, validation, hyperparameter tuning) as required by the inclusion criteria, or where the data partitioning strategy introduces a high risk of data leakage (e.g., patient-wise contamination between sets).
(4) Studies where the reported outcome data (e.g., contingency table, sensitivity, specificity, AUC) are derived from a pooled analysis that includes the training or validation sets, rather than exclusively from the independent test set.
(5) Studies published in a format that is not an original, peer-reviewed research article (e.g., conference abstracts without a subsequent full paper, editorials, commentaries, study protocols, or theses).

Information sources Web of Science, PubMed, EMBASE and Cochrane Library.

Main outcome(s) Primary outcomes: Pooled sensitivity, specificity, and SROC AUC for AI-ECG prediction of AF, estimated by bivariate random-effects meta-analysis.Secondary outcomes: (1) Diagnostic accuracy stratified by prediction horizon (short/mid/long-term); (2) Pooled AUC difference between AI-ECG and CHARGE-AF score; (3) Hazard ratios for incident AF by AI-predicted risk.
Timing: Prediction windows ranging from ≤ 1 month to 5 years.
Effect measures: Sensitivity, specificity, AUC, AUC difference, hazard ratios; all reported with 95% confidence intervals. Heterogeneity assessed by I^2 and τ(tau).

Quality assessment / Risk of bias analysis Risk of bias was assessed with QUADAS-2 across four domains (patient selection, index test, reference standard, flow and timing). Selective reporting was also noted. Two reviewers independently rated each study as low, high, or unclear risk; disagreements were resolved by consensus or a third reviewer.

Strategy of data synthesis A bivariate random-effects model (Reitsma method) was used to pool sensitivity and specificity, with hierarchical SROC curves generated. Heterogeneity was assessed with I^2 and τ. Subgroup analyses were performed by prediction horizon (short-term, mid-term, and long-term). Meta-regression examined the effects of prediction horizon and clinical scenario on sensitivity. For studies reporting both AI-ECG and CHARGE-AF AUCs, a fixed-effect meta-analysis of AUC differences was conducted. Hazard ratios for incident AF by AI-predicted risk were summarised descriptively due to definitional heterogeneity. Publication bias was evaluated with Egger's test and the trim-and-fill method. All analyses were performed in R 4.3.2 using the mada, meta, and metafor packages.

Subgroup analysis Subgroup analyses were performed by prediction horizon: short-term (≤1 month), mid-term (1 month to 1 year), and long-term (>1 year).

Sensitivity analysis Sensitivity of the results to publication bias was assessed using Egger's test and the trim-and-fill method.

Language restriction English.

Country(ies) involved China.

Keywords Atrial fibrillation; Artificial intelligence; Electrocardiogram; Diagnostic accuracy; Meta-analysis.

Contributions of each author

Author 1 - HUANG Yexin.
Email: 18039058565@163.com
Author 2 - Zhang Xiling.
Email: zxll1027@163.com
Author 3 - WANG Xinkang.
Email: 13305027368@163.com