

Artificial intelligence for arrhythmia detection and prediction

Edited by

Evangelos Oikonomou, Polydoros Kampaktsis and
Alexios Antonopoulos

Coordinated by

Panteleimon Pantelidis

Published in

Frontiers in Cardiovascular Medicine
Frontiers in Artificial Intelligence



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-7177-4
DOI 10.3389/978-2-8325-7177-4

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Artificial intelligence for arrhythmia detection and prediction

Topic editors

Evangelos Oikonomou — National and Kapodistrian University of Athens, Greece

Polydoros Kampaktsis — Columbia University Irving Medical Center, Columbia University, United States

Alexios Antonopoulos — National and Kapodistrian University of Athens, Greece

Topic coordinator

Panteleimon Pantelidis — National and Kapodistrian University of Athens, Greece

Citation

Oikonomou, E., Kampaktsis, P., Antonopoulos, A., Pantelidis, P., eds. (2025). *Artificial intelligence for arrhythmia detection and prediction*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7177-4

Table of contents

- 05 **Editorial: Artificial intelligence for arrhythmia detection and prediction**
Panteleimon Pantelidis, Alexios Antonopoulos, Polydoros Kampaktis and Evangelos Oikonomou
- 08 **Artificial intelligence-enabled atrial fibrillation detection using smartwatches: current status and future perspectives**
Zoi Papalamprakopoulou, Dimitrios Stavropoulos, Serafeim Moustakidis, Dimitrios Avgerinos, Michael Efremidis and Polydoros N. Kampaktis
- 16 **AI-enabled detection of QRS fragmentation from 12-lead electrocardiogram and its clinical relevance for predicting malignant arrhythmia onset**
Sebastian Ingelaere, Amalia Villa, Carolina Varon, Sabine Van Huffel, Bert Vandenberg and Rik Willems
- 25 **A new prediction model for sustained ventricular tachycardia in arrhythmogenic cardiomyopathy**
Baowei Zhang, Xin Xie, Jinbo Yu, Yizhang Wu, Jian Zhou, Xiaorong Li and Bing Yang
- 34 **Cardioattentionnet: advancing ECG beat characterization with a high-accuracy and portable deep learning model**
Youfu He, Yu Zhou, Yu Qian, Jingjie Liu, Jinyan Zhang, Debin Liu and Qiang Wu
- 46 **Artificial intelligence enabled mobile health technologies in arrhythmias-an opinion article on recent findings**
Arijita Banerjee
- 51 **Bioinformatics analysis of the expression of potential common genes and immune-related genes between atrial fibrillation and chronic kidney disease**
Jieying Teng and Guoxiong Deng
- 69 **Deep learning analysis of exercise stress electrocardiography for identification of significant coronary artery disease**
Hsin-Yueh Liang, Kai-Cheng Hsu, Shang-Yu Chien, Chen-Yu Yeh, Ting-Hsuan Sun, Meng-Hsuan Liu and Kee Koon Ng
- 81 **Identification of atrial fibrillation using heart rate variability: a meta-analysis**
Ziwei Yin, Changxin Liu, Chenggong Xie, Zixing Nie, Jiaming Wei, Wen Zhang and Hao Liang

- 91 **Artificial intelligence in atrial fibrillation: emerging applications, research directions and ethical considerations**
Ibrahim Antoun, Ahmed Abdelrazik, Mahmoud Eldesouky, Xin Li, Georgia R. Layton, Mustafa Zakkar, Riyaz Somani and G. André Ng
- 104 **ECG-XPLAIM: eXplainable Locally-adaptive Artificial Intelligence Model for arrhythmia detection from large-scale electrocardiogram data**
Panteleimon Pantelidis, Samuel Ruipérez-Campillo, Julia E. Vogt, Alexios Antonopoulos, Ioannis Gialamas, George E. Zakyntinos, Michael Spartalis, Polychronis Dilaveris, Jose Millet, Theodore G. Papaioannou, Evangelos Oikonomou and Gerasimos Siasos



OPEN ACCESS

EDITED AND REVIEWED BY
Rui Providencia,
University College London, United Kingdom

*CORRESPONDENCE
Panteleimon Pantelidis
✉ ppantele@med.uoa.gr

RECEIVED 18 October 2025
ACCEPTED 27 October 2025
PUBLISHED 06 November 2025

CITATION
Pantelidis P, Antonopoulos A, Kampaktis P
and Oikonomou E (2025) Editorial: Artificial
intelligence for arrhythmia detection and
prediction.
Front. Cardiovasc. Med. 12:1727758.
doi: 10.3389/fcvm.2025.1727758

COPYRIGHT
© 2025 Pantelidis, Antonopoulos, Kampaktis
and Oikonomou. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License (CC
BY). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Editorial: Artificial intelligence for arrhythmia detection and prediction

Panteleimon Pantelidis^{1*}, Alexios Antonopoulos²,
Polydoros Kampaktis³ and Evangelos Oikonomou¹

¹3rd University Department of Cardiology, National and Kapodistrian University of Athens, Athens, Greece, ²1st University Department of Cardiology, National and Kapodistrian University of Athens, Athens, Greece, ³Division of Cardiology, Columbia University Irving Medical Center, New York, NY, United States

KEYWORDS

arrhythmia, diagnostics, deep learning, artificial intelligence, risk prediction

Editorial on the Research Topic

Artificial intelligence for arrhythmia detection and prediction

The convergence of artificial intelligence (AI) and cardiovascular medicine has reached a critical juncture, offering unprecedented opportunities that can revolutionize arrhythmia care. This Research Topic presents ten pioneering articles that span from advanced detection algorithms to emerging prediction models, charting a clear trajectory toward truly predictive, personalized applications in cardiac arrhythmias.

1 Advanced detection technologies: achieving clinical excellence

He et al. introduce CardioAttentionNet (CANet), a portable deep learning (DL) model combining advanced architectures to classify five arrhythmia types with accuracies around: $97.37\% \pm 0.30\%$ (normal), $98.09\% \pm 0.25\%$ (supraventricular tachycardia), $92.92\% \pm 0.09\%$ (ventricular tachycardia), $99.07\% \pm 0.13\%$ (atrial fibrillation, AF), and $99.68\% \pm 0.06\%$ (unclassified arrhythmias). AUCs exceed 99%, with 94.41% average accuracy on external data and 56.7 ms inference per electrocardiogram (ECG), demonstrating real-time feasibility. Complementing this, ECG-XPLAIM delivers a novel, explainable, inception-style model with multi-scale kernels (4, 20, 80 ms). Internal validation on MIMIC-IV and external on PTB-XL achieved AUCs of 0.88–0.99 and overall sensitivities/specificities >0.9 . Grad-CAM visualizations offer insights into the model's predictions, bridging the gap between AI performance and clinical interpretability (Pantelidis et al.).

Summarizing relevant data, Yin et al.'s meta-analysis of heart rate variability (HRV)-based AF detection demonstrates a pooled sensitivity of 94% and specificity of 97%, with DL outperforming traditional machine learning, supporting AI readiness across diverse cohorts. Similarly, Papalamprakopoulou et al. review AI-enabled smartwatch AF detection, illustrating how continuous, noninvasive monitoring by wearable devices enhances preventive cardiology, while underscoring challenges in reducing false positives and integrating vast data streams into clinical workflows.

2 Emerging prediction capabilities: from risk assessment to personalized forecasting

Beyond accurate detection, contemporary AI is pivoting toward the more ambitious objective of forecasting arrhythmias and quantifying individualized risk. Zhang et al. propose a nomogram for sustained ventricular tachycardia in arrhythmogenic cardiomyopathy integrating age, sex, syncope, heart failure, echocardiographic parameters, and ECG findings. AUCs of 0.867 (training) and 0.815 (validation) mark significant strides in personalized risk stratification. Ingelaere et al. assess AI-enabled QRS fragmentation detection from 12-lead ECGs in 1,242 patients with implantable cardioverter-defibrillator (ICD) using support vector machines and features from phase-rectified signal averaging and variational mode decomposition. Despite a sensitivity of 0.76 and specificity of 0.92 for fragmentation, no significant link to ICD therapy or mortality was found, revealing limitations when translating traditional ECG markers into AI-based risk predictors. Liang et al.'s neural network for exercise stress ECG achieved AUC 0.83, sensitivity 0.89, and specificity 0.60 for significant coronary artery disease (CAD) detection, processing ECG segments, and clinical features. Results are delivered within one minute post-test, spotlighting the potential to streamline diagnostic workflows.

3 More clinical, molecular and multimodal applications

Antoun et al. provide a comprehensive scoping review of AI in AF across the full care continuum, emphasizing its value for identifying asymptomatic AF, tailoring therapy, and confronting key implementation barriers such as transparency, data integration, and regulatory alignment. Banerjee discusses mobile health AI technologies, detailing smartphone-based ECG and photoplethysmography (PPG) systems that democratize long-term rhythm monitoring, especially in underserved regions, and emphasizing accessibility alongside diagnostic accuracy. From a molecular viewpoint, Teng and Deng explore shared biology between AF and chronic kidney disease, identifying co-expressed genes (PPBP, CXCL1, RSAD2) and transcription factors (FOXC1, FOXL1, GATA2); this bioinformatics pipeline illustrates how AI can reveal mechanistic targets and candidate biomarkers, advancing precision therapeutics and risk stratification at the genotype-phenotype interface.

4 Current landscape and future directions

AI-driven detection systems for arrhythmias can now rival expert cardiologists. CANet and ECG-XPLAIM report sensitivities and specificities above 94%, and short inference times per ECG segment, enabling real-time deployment on portable

and wearable devices. A meta-analysis of DL, HRV-based AF detection confirms the impressive performance across diverse cohorts. These advances support population-level screening via smart devices, lowering barriers to early AF identification (1).

By contrast, AI-based prediction models remain an emerging frontier. Current algorithms for ventricular tachycardia risk and post-exercise CAD detection achieve AUCs of 0.81–0.87. These models integrate clinical and lab test parameters, yet seem to lack robust external validation. Similarly, AI-detected QRS fragmentation shows reasonable detection accuracy, but no correlation with ICD therapy or mortality, highlighting the challenge of converting traditional ECG markers into prognostic tools. To achieve clinical maturity, prediction models could be enhanced by integrating multimodal data across large, diverse cohorts with longitudinal follow-up (2). Additionally, novel network architectures further drive advances in cardiac AI: Self-attention, applied to sequential ECG data, can capture long-range temporal dependencies and spotlight clinically salient signal segments. Even the more traditional, inception-style designs seem to work exceptionally well, enabling multiscale feature extraction and enhancing detection robustness. Depthwise separable convolutions and MobileNet-inspired modules ensure computational efficiency on constrained hardware (Pantelidis et al., Banerjee, Teng and Deng, 3).

However, critical hurdles remain. Explainability is essential for clinician trust and regulatory approval; Grad-CAM visualizations are increasingly integrated to reveal model decision pathways. Data heterogeneity, concerning varying ECG hardware, sampling rates, and annotation standards, limits generalizability and demands rigorous standardization. Interoperability across electronic health records and wearable platforms requires harmonized data formats. Moreover, evolving regulatory frameworks for AI as a medical device necessitate transparent validation, real-world performance monitoring, and post-market surveillance to detect algorithmic drift and bias (3–5).

The next frontier lies in multimodal integration: combining continuous ECG, imaging, genomic, and environmental data for comprehensive risk modeling. Agentic AI could autonomously adjust monitoring and therapeutic recommendations in real time, optimizing care pathways. Transformer-based architectures tailored to multidimensional cardiac data may predict imminent arrhythmic events by detecting subtle premonitory signal shifts. Digital twin frameworks, enabling virtual patient models continuously updated with live data, could simulate arrhythmic risk under different interventions, enhancing personalized preventive strategies (6). Realizing predictive, personalized arrhythmia care will require sustained multidisciplinary collaboration among cardiologists, data scientists, bioengineers, and ethicists. By addressing explainability, data integration, and equitable access, AI-enabled prediction can transition from research prototypes to life-saving clinical tools, heralding a new era of proactive cardiovascular medicine.

Author contributions

PP: Conceptualization, Formal analysis, Methodology, Writing – original draft. AA: Methodology, Supervision,

Writing – review & editing. PK: Methodology, Supervision, Writing – review & editing. EO: Methodology, Supervision, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Perone F, Bernardi M, Redheuil A, Mafrica D, Conte E, Spadafora L, et al. Role of cardiovascular imaging in risk assessment: recent advances, gaps in evidence, and future directions. *J Clin Med.* (2023) 12:5563. doi: 10.3390/jcm12175563
2. Escaned J, Berry C, De Bruyne B, Shabbir A, Collet C, Lee JM, et al. Applied coronary physiology for planning and guidance of percutaneous coronary interventions. A clinical consensus statement from the European Association of Percutaneous Cardiovascular Interventions (EAPCI) of the European Society of Cardiology. *EuroIntervention.* (2023) 19:464–81. doi: 10.4244/EIJ-D-23-00194
3. Khera R, Oikonomou EK, Nadkarni GN, Morley JR, Wiens J, Butte AJ, et al. Transforming cardiovascular care with artificial intelligence: from discovery to practice. *J Am Coll Cardiol.* (2024) 84:97–114. doi: 10.1016/j.jacc.2024.05.003
4. Pantelidis P, Dilaveris P, Ruipérez-Campillo S, Goliopoulou A, Giannakodimos A, Theofilis P, et al. Hearts, data, and artificial intelligence wizardry: from imitation to innovation in cardiovascular care. *Biomedicines.* (2025) 13:1019. doi: 10.3390/biomedicines13051019
5. Vandenberk B, Chew DS, Prasana D, Gupta S, Exner DV. Successes and challenges of artificial intelligence in cardiology. *Front Digit Health.* (2023) 5:1201392. doi: 10.3389/fdgth.2023.1201392
6. Pantelidis P, Oikonomou E, Gialamas I, Goliopoulou A, Sarantos S, Zakynthinos GE, et al. Decoding the heart: how artificial intelligence is transforming cardiology. *J Med Artif Intell.* (2025) 8:9–9. doi: 10.21037/jmai-24-139



OPEN ACCESS

EDITED BY

Patrick Badertscher,
University Hospital of Basel, Switzerland

REVIEWED BY

Vassilios Vassilikos,
Aristotle University of Thessaloniki, Greece

*CORRESPONDENCE

Polydoros N. Kampaktsis
✉ pkampaktsis@yahoo.com

RECEIVED 14 May 2024

ACCEPTED 02 July 2024

PUBLISHED 15 July 2024

CITATION

Papalamprakopoulou Z, Stavropoulos D,
Moustakidis S, Avgerinos D, Efremidis M and
Kampaktsis PN (2024) Artificial intelligence-
enabled atrial fibrillation detection using
smartwatches: current status and future
perspectives.
Front. Cardiovasc. Med. 11:1432876.
doi: 10.3389/fcvm.2024.1432876

COPYRIGHT

© 2024 Papalamprakopoulou, Stavropoulos,
Moustakidis, Avgerinos, Efremidis and
Kampaktsis. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Artificial intelligence-enabled atrial fibrillation detection using smartwatches: current status and future perspectives

Zoi Papalamprakopoulou¹, Dimitrios Stavropoulos²,
Serafeim Moustakidis³, Dimitrios Avgerinos⁴, Michael Efremidis⁴
and Polydoros N. Kampaktsis^{5*}

¹Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY, United States, ²Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States, ³AiDEAS, Tallinn, Estonia, ⁴Onassis Cardiac Surgery Center, Athens, Greece, ⁵Department of Medicine, Aristotle University of Thessaloniki Medical School, Thessaloniki, Greece

Atrial fibrillation (AF) significantly increases the risk of stroke and heart failure, but is frequently asymptomatic and intermittent; therefore, its timely diagnosis poses challenges. Early detection in selected patients may aid in stroke prevention and mitigate structural heart complications through prompt intervention. Smartwatches, coupled with powerful artificial intelligence (AI)-enabled algorithms, offer a promising tool for early detection due to their widespread use, easiness of use, and potential cost-effectiveness. Commercially available smartwatches have gained clearance from the FDA to detect AF and are becoming increasingly popular. Despite their promise, the evolving landscape of AI-enabled smartwatch-based AF detection raises questions about the clinical value of this technology. Following the ongoing digital transformation of healthcare, clinicians should familiarize themselves with how AI-enabled smartwatches function in AF detection and navigate their role in clinical settings to deliver optimal patient care. In this review, we provide a concise overview of the characteristics of AI-enabled smartwatch algorithms, their diagnostic performance, clinical value, limitations, and discuss future perspectives in AF diagnosis.

KEYWORDS

atrial fibrillation, smartwatches, photoplethysmography, single-lead electrocardiography, artificial intelligence, machine learning

Introduction

Atrial fibrillation (AF) constitutes the most common cardiac arrhythmia with a prevalence that is increasing. According to the Global Burden of Diseases study, the number of individuals affected by AF has doubled from 1990 to 2019, reaching approximately 60 million globally (1). This upward trajectory is projected to continue, with estimates suggesting that by 2030, the United States alone could see over 12.1 million people diagnosed with AF (2).

AF is associated with a fivefold increased risk of stroke, as well as an increased risk for heart failure (3, 4). However, its episodic manifestation and frequently asymptomatic nature, pose significant challenges for AF diagnosis (5). For instance, research data suggest that paroxysmal AF can be detected in 10%–20% of patients with cryptogenic

strokes and extended monitoring with implantable devices is recommended in this population to increase sensitivity of AF detection (6–8). In selected asymptomatic patients, early AF detection could be helpful in preventing stroke using oral anticoagulation (OAC) or preventing heart failure using rhythm control strategies (4, 9).

Smartwatches empowered by artificial intelligence (AI) algorithms have emerged as a promising tool for early detection of AF (10, 11). Recent cost-effectiveness studies also underscore their potential to serve as a widespread tool for early arrhythmic detection (12). However, AF confirmation still requires electrocardiography (ECG) and widespread incorporation of smartwatches for AF detection has not been established (13–15).

In this review we first summarize the characteristics and performance of smartwatches that use AI-enabled algorithms for AF detection; secondly, we discuss their clinical applicability, the challenges around their use, and finally outline future directions in the field.

Smartwatch technology for AF detection

Photoplethysmography

A common technology for detecting AF with smartwatches is photoplethysmography (PPG). This technique involves illuminating the skin with a light-emitting diode (LED) and detecting the light reflected back with a photodetector (16). The intensity of the reflected light constitutes the PPG signal, which varies according to blood volume changes in the vessels throughout the cardiac cycle. Therefore, the PPG signal represents a pulse pressure waveform that enables passive, continuous, or semi-continuous heart rhythm monitoring measured at the wrist (17).

Since various neural, cardiac, and respiratory factors regulate blood flow, physiologic cardiovascular parameters such as heart rate, blood pressure, oxygen saturation, and respiratory rate could be derived from PPG signal analysis (18). Importantly, the peak of the PPG signal correlates closely with the R wave of the electrocardiogram (ECG) (19). As such, the PPG signal can provide R-R intervals, i.e., heart rate variability (20) and in AF, the PPG signal exhibits greater irregularity of pulse waveform (16). Assessment of PPG variability through a detection algorithm is used to detect AF.

Single-lead ECG

Another common method for detecting AF with smartwatches is based on the recording of a single-lead ECG. In this method, the watch's back acts as a positive electrode and the contralateral fingertip is placed on the crown, acting as a negative electrode (21). This creates a bipolar ECG-lead, simulating Einthoven's ECG lead I. The recorded ECG tracing is typically saved and can be analyzed either manually by physicians or automatically. In addition to single-lead ECG, other techniques further explore the use of a smartwatch-based wireless 6-lead limb-like ECG, which requires interpretation by cardiologists, combined with PPG to increase accuracy of AF detection (22).

Detection algorithms and machine learning

Detection algorithms may rely on traditional statistics or AI, particularly machine learning (ML) and deep learning (DL) algorithms (23). For commercially available smartwatches, the AI-empowered algorithms utilized for AF detection remain proprietary to each company (24, 25). Nonetheless, we will provide a brief overview of how similar algorithms described in the research setting function to aid physicians in understanding their mechanisms. PPG or ECG signals, once captured, undergo several preprocessing steps including noise reduction, normalization, and segmentation. Feature extraction focuses on both time-domain and frequency-domain attributes, which might include heart rate variability indices, the morphology of the PPG or ECG waveform, and temporal intervals between heartbeats. A tachogram, a visual representation of heart rate variability plotted against time, serves as a key component in the PPG-based AF detection process (26). Typically, a set number of irregular tachograms over a period of time is involved in the algorithm's decision to trigger an irregular pulse notification (IPN). ML models such as support vector machines and random forests analyze these engineered features to classify heart rhythms, benefitting from the clear delineation of feature-based input (27, 28). In contrast, DL models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can process raw or minimally preprocessed signal data (29, 30). These models are adept at automatically detecting complex patterns within large datasets, which is beneficial for identifying subtle and non-linear indicators of AF. CNNs, for instance, are useful for spatial pattern recognition within ECG signals, while RNNs excel at analyzing sequential data, capturing dynamic changes over time which are crucial for continuous ECG monitoring (31).

Challenges

PPG sensors can yield missing signals or inconclusive/unclassified rhythms. Missing signals are often due to poor sensor-skin contact. On the other hand, insufficient quality of the signal may be attributed to motion artifacts, such as from muscular motion. Overall, PPG sensors tend to underestimate AF detection at both higher and lower heart rates when compared to standard detection methods (32). Additionally, ectopic beats can further complicate AF detection by generating pulse irregularities (22). Although evidence remain inconclusive, skin pigmentation may also restrict PPG performance (33, 34). Moreover, the PPG signal may be weaker in elderly, due to pathologic or physiologic changes associated with aging including reduced peripheral blood flow, increased arterial wall stiffness, and skin changes (35, 36). On the other hand, there may be a low amplitude of P-waves with single-lead ECG, which challenges the classification of electrical activity (37). Lastly, limitations of using smartwatch to detect AF include the necessity for user cooperation and regular charging, depending on its battery life (10).

Similarly, AI models, despite their capabilities, require robust, diverse datasets for training to ensure effective performance across various demographic groups and conditions. Additionally, these models necessitate significant computational resources,

particularly in DL frameworks, to process extensive data and learn through sophisticated layers of abstraction. The interpretability of these AI systems is also crucial, particularly in healthcare settings, where providers prefer algorithms that offer insights into their decision-making processes (38). This aligns with the increasing demand for explainable AI in clinical environments, highlighting another significant limitation.

Diagnostic performance of Ai-empowered smartwatches in AF detection

Currently, AF diagnosis requires that a physician reviews a standard 12-lead ECG or rhythm strips (9, 39). This also serves as the gold standard to determine the diagnostic accuracy of smartwatches. Table 1 summarizes key recent studies conducted from 2019 and onwards that evaluated the diagnostic performance of smartwatches for detecting AF.

Large-scale, pragmatic studies that were conducted for major smartwatch companies like Apple and Fitbit assessed the ability to screen for AF in the general population relying on PPG (40, 41). The Apple Heart Study evaluated over 400,000 individuals (mean age 41 ± 13 years, 42% women) for IPN with the Apple Watch indicating possible AF in that event (40). The duration of PPG monitoring and the criteria to triggering an IPN for AF varied significantly among studies assessing PPG sensors. PPG monitoring was more intermittent with the Apple algorithm, where the IPN required at least 5/6 consecutive irregular 1-minute tachograms within a 48-hour period (40). Conversely, the Fitbit algorithm allowed for more continuous monitoring and applied stricter notification criteria, requiring at least 30 min of irregular rhythm before notifying the user (41). Participants who received such an IPN notification received a single-lead ECG patch via mail to wear for 1 week, which was used to diagnose AF. Overall, only 0.52% of the study population received an IPN using the Apple algorithm, with the percentage correlating with increased age (≥ 65 years: IPN 3.2%). Similarly, the Fitbit Heart Study assessed the performance of the Fitbit Watch among over 400,000 participants (median age 47 years, interquartile range 35–58 years, 71% women), of whom 1% received an IPN (≥ 65 years: IPN 3.6%) (41). Of those who received an IPN in the Fitbit and Apple Heart studies, only up to 25% of all notified participants in these studies returned the ECG patch. 32.2% and 34.0%, respectively were confirmed to have AF lasting at least 30 s on the reference ECG patch. The positive predictive value (PPV) for AF, confirmed concurrently on the ECG patch, of the Apple algorithm was 84.0% (95% CI, 76.0%–92.0%). The PPV was lower among those over 65, i.e., 78% (95% CI, 64.0%–92.0%). The Fitbit algorithm yielded a sensitivity of 67.6%, specificity of 98.4%, and PPV of 98.2% (95% CI, 95.5%–99.5%), with a slight reduction among those aged ≥ 65 years at 97.0% (95% CI, 91.4%–99.4%).

Among studies conducted in research settings and among populations at high-risk for AF, the performance of PPG sensors varied with sensitivity ranging from 87.8% (42) to 94.2% (22) and specificity up to 99.1% (22). Other studies examined the performance of intelligent ECG (iECG), a smartwatch-based

single-lead ECG with automatic AF detection function (43–46). AF was considered present if episodes lasted for at least 30 s. The sensitivity of automatic AF detection with iECG varied from as low as 41.0% among post-cardiac surgery patients (44) to 94.4% among high-risk inpatient populations (45), while the specificity ranged from 69.0% (46) to 100.0% (44). These performance indices should be interpreted with caution, as most of these studies employed intention-to-diagnose analyses, excluding a significant number of tracings that were deemed inconclusive. Badertscher et al. demonstrated a reduction in such inconclusive tracings from 14% to 4.1% after a cardiologist's review, and this was further supported by Mannhart et al. (43, 46). Avram et al. demonstrated that obtaining a confirmatory iECG one hour apart the first one improved the specificity (100.0%) without affecting sensitivity (96.0%). Moreover, novel AI algorithms have demonstrated excellent performance in reducing inconclusive classification rates (47). Lastly, in most studies, patients were instructed on how to use the technology for AF detection which may not reflect real-life circumstances or digital literacy, especially among the elderly.

Clinical impact of AF detection using Ai-empowered smartwatches

Potential benefits

The main potential advantage of detecting asymptomatic AF using smartwatches lies in the opportunity for early diagnosis and therapeutic intervention. Identifying AF among high risk, asymptomatic patients could lead to the initiation of OAC, potentially reducing the risk for stroke (4). The eBRAVE-AF trial showed that PPG-based screening more than doubled the detection rate of asymptomatic AF leading to subsequent OAC initiation (48). A recent meta-analysis demonstrated a significant reduction in stroke risk after OAC initiation in patients with asymptomatic AF detected by implantable cardiac devices (49). However, the results of studies investigating the specific clinical benefits of smartwatch-detected AF and the populations to which may apply, are pending (50). Additionally, increasing evidence suggests that early initiation of rhythm control not only lowers the risk of stroke, but also reduces the structural, long-term complications of AF, particularly HF (51, 52). Furthermore, identifying AF early could prompt the early identification and management of other concomitant cardiovascular conditions such as hypertension, underlying heart disease and diabetes mellitus (53). Lastly, population screening for AF using smartwatches is proposed as a more cost-effective approach compared to no screening or conventional screening (12). Nevertheless, no consensus has been reached yet regarding AF screening due to a lack of clinical data to support its potential benefits in asymptomatic patient groups at risk for AF and its complications (5, 9, 54).

Identifying populations for smartwatch-based AF detection

In the absence of relevant guidelines, it may be reasonable to introduce smartwatch AF screening in certain populations. For

TABLE 1 Key studies that evaluated the performance of smartwatches for AF detection.

Authors (year)	Study type	Aim of the study	Number of participants (N)	Study population	Smartwatch	Sensor/confirmatory test	Reference standard	Incidence & diagnostic performance
Perez et al. (40)	Prospective, siteless, pragmatic, population study	Screen for AF with IPN and confirm diagnosis with continuous ECG monitoring	419,297	≥22 yo with a compatible Apple iPhone and Apple Watch 6% were aged ≥65 yo Exclusions: known AF, OACs	Apple Watch Series 1–3	PPG	Single-lead ECG patch monitoring ×1 week in those with IPN only (clinician-interpreted)	IPN = 0.52% PPV = 84.0%
Lubitz et al. (41)	Prospective, siteless, pragmatic, population study	Screen for AF with IPN and confirm diagnosis with continuous ECG monitoring	455,699	≥22 yo with a compatible Fitbit wrist-worn device and Android/iOS smartphones 12% were aged ≥65 yo Exclusions: known AF, OACs, pacemaker/defibrillator	Fitbit wrist-worn device	PPG	Single-lead ECG patch monitoring ×1 week in those with IPN only (clinician-interpreted)	IPN = 1.0% PPV = 98.2% SN = 67.6% SP = 98.4%
Badertscher et al. (43)	Prospective, observational, single-center study	Assess the diagnostic accuracy of a novel smart device in detecting AF in cardiac patients	319	Patients who presented to the cardiology department Median = 67 yo; interquartile range 54–76 yo.	Withings Scanwatch	Single-lead iECG	12-lead ECG (cardiologist-interpreted)	SN = 76.0% SP = 99.0% $\kappa = 0.72$
Bacevicius et al. (22)	Single-center, non-randomized validation study with a prospective case-control model	Assess the diagnostic accuracy of a novel system in detecting AF when challenged by PVCs/PACs	344 (AF: n = 121)	≥18 yo, known AF, SR, or SR with frequent PVCs/PACs Exclusions: regular pulse wave despite AF, other arrhythmia	Wrist-worn device	PPG/on-demand watch 6-lead ECG	ECG Holter monitor (cardiologist-interpreted)	PPG + 6-lead ECG for the PVCs/PACs group: SN = 94.2%, SP = 99.6%, Accuracy = 99.5%
Avram et al. (42)	Prospective remote-based study	Estimate the SN and SP of a smartwatch for continuous detection of AF and calculate AF burden	204	Known AF or at risk for AF (age ≥ 65 yo, history of hypertension, diabetes, heart failure, or coronary disease) 48.0% were aged ≥65 yo	Samsung Galaxy Watch Active 2	PPG/on-demand watch single-lead iECG	Continuous 28-day ECG patch monitor (cardiologist and technician-interpreted)	PPG; SN = 87.8% SP = 97.4% PPG + single-lead ECG; SN = 96.9 SP = 99.3
Seshadri et al. (44)	Prospective	Assess the diagnostic accuracy of a smartwatch in detecting AF in cardiac patients	50	Patients who had undergone cardiac surgery and were on telemetry Exclusions: pacemaker, surgical use of a radial artery, chronic HR abnormalities, wrist tattoos	Apple Watch Series 4 (AW4) that used a notification and a downloadable waveform on PDF	Single-lead iECG	Telemetry	AW4 notification: SN = 41.0% SP = 100.0%
Rajakariar et al. (45)	Prospective, multi-center, validation study	Evaluate the accuracy of a novel smartwatch-based ECG monitor for diagnosing AF in an unselected high-risk inpatient population	200 patients (AF: n = 38)	Consecutive patients ≥18 yo admitted to the medical, cardiac or intensive care ward Exclusions: cardiac implantable devices, unable to independently use the device, or in contact isolation	AliveCor KardiaBand smartwatch-based HR monitor	Single-lead iECG	12-lead ECG (cardiologist-interpreted)	SN = 94.4% SP = 81.9% PPV = 54.8% NNP = 98.4% $\kappa = 0.60$

(Continued)

TABLE 1 Continued

Authors (year)	Study type	Aim of the study	Number of participants (N)	Study population	Smartwatch	Sensor/confirmatory test	Reference standard	Incidence & diagnostic performance
Mannhart et al. (46)	Prospective, single-center, diagnostic study	Assess the accuracy of 5 smart devices in identifying AF	201 patients (AF: n = 62)	≥18 yo scheduled for catheter ablation procedures, electric cardioversions, pacemaker, or ICD implantation Exclusions: 3 or less single-lead ECG recordings due to logistical or technical issues or missing data.	Apple Watch 6, AliveCor KardiaMobile 6l, KardiaMobile or KardiaMobile 6l, Fitbit Sense, Samsung Galaxy Watch 3, and Withings Scanwatch ECG	Single-lead iECG	12-lead ECG (physician-interpreted)	Apple Watch: SN = 85.0%, SP = 75.0% Samsung Galaxy Watch 3: SN = 85.0% SP = 75.0% Withings Scanwatch: SN = 58.0% SP = 75.0% Fitbit Sense: SN = 66.0% SP = 79.0% AliveCor KardiaMobile: SN = 79.0% SP = 69.0%

AF, atrial fibrillation; IPN, irregular pulse notification; ECG, electrocardiogram; yo, years old; iECG, intelligent electrocardiogram; OACs, oral anticoagulants; PPG, photoplethysmography; PPV, positive predictive value; SN, sensitivity; SP, specificity; CI, confidence interval; κ , kappa coefficient; HR, heart rhythm; PVCs, premature ventricular contractions; PACs, premature atrial contractions; SR, sinus rhythm; ICD, implantable cardioverter-defibrillator; TIA, transient ischemic attack.

example, the 2020 ESC Guidelines for AF recommend AF screening for patients over 75 (5). Additionally, the same guidelines suggest screening for those at high stroke risk, though the optimal cut-offs on assessment tools, e.g., CHA2DS2-VASc score, for smartwatch-based AF detection require further clarification. Opportunistic screening for AF is also suggested for patients aged 65 years and above, where smartwatches have already been incorporated as screening options (5). In contrast, the US Preventive Services Task Force implies that the current evidence is insufficient to assess the balance of benefits and harms for AF screening (54). It is noteworthy that AI algorithms utilizing Electronic Medical Records (PULsE-AI, FIND-AF) or ECGs (BEAGLE) may facilitate the identification of high-risk patients (55–57). However, it would be difficult to envision that smartwatches could be used for screening or substitute other established methods of AF detection in symptomatic patients at high risk for AF, such as those with cryptogenic stroke.

Clinical challenges and risks

Once the smartwatch indicates possible AF, the main question is how healthcare providers should proceed (10). In existing clinical studies, individuals were referred to contact a healthcare provider (40, 41, 58). Healthcare providers then confirmed the presence of AF through a standard ECG confirmatory test. This approach aligns with the 2020 ESC Guidelines for AF, which included smartwatches in the screening recommendations (5). However, the introduction of smartwatches for AF diagnosis poses risks for putting high stress on the healthcare system, particularly if extended population screening is implemented (59, 60). Therefore, it is crucial to optimize the parameters of smartwatch algorithms to minimize false positives and prevent physician overload. A study conducted among patients with known AF who utilized wearables, demonstrated increased engagement in follow-up healthcare, with no significant difference in pulse rates compared to those not using wearables (61). Further research into health outcomes will guide management for both patients and providers.

Data management

Smartwatch data management poses an ongoing challenge. The FDA does not classify smartwatches as medical devices; rather they are considered wellness tools, subject to expedited regulatory approval through the Digital Health Software Precertification (Pre-Cert) Program (62). This regulatory framework promotes the development of a large digital health market and the generation of a substantial volume of digital data (21). Despite this, smartwatch data have not yet been fully integrated into clinical workflows. While meaningful integration of smartwatch data into the existing Electronic Health Records (EHRs) will eventually become necessary, the specific type and volume of data to be shared with EHRs is yet to be determined. Consequently, there is an ongoing effort to endorse partnerships between startup organizations that specialize in integrating such data into EHRs with healthcare systems and insurance companies (63). It is also crucial to develop clear

legislation regarding data ownership, ensure compliance with HIPAA regulations and address patient confidentiality, privacy, and security concerns (59, 64). Furthermore, attention is required concerning physician reimbursement, including the creation of standard documentation processes and billable codes related to the integration of smartwatch data into clinical workflows (65). With all these impending changes, it has been suggested that a “digital health counselor” could provide support and guide patients and providers throughout the initial transition phase of digital data integration into clinical workflows (59).

Conclusions and future directions

AI-empowered smartwatches have emerged as a potential screening tool for AF, the most common arrhythmia with an increasing prevalence. Potentially important advantages of this strategy include purely non-invasive nature, easiness of use, early detection of AF in high-risk asymptomatic individuals and reduction in AF burden. However, there is no consensus among pertinent societies regarding AF screening and the clinical impact of this strategy using smartwatches remains to be proven. Particular issues that have to be addressed are: identification of populations that could benefit from AF screening using smartwatches; optimization of AI algorithms to improve diagnostic accuracy and reduce confirmatory tests; data management of protected health care information including infrastructure for processing of large amounts of data, and clinical algorithms to assist with widespread application of AF screening via smartwatches. Large trials to address these questions are currently required and may expect the landscape of AF in the near future.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet*. (2020) 396(10258):1204–22. doi: 10.1016/S0140-6736(20)30925-9
2. Prevention CfDCA. Atrial Fibrillation. (2022). Available online at: https://www.cdc.gov/heartdisease/atrial_fibrillation.htm (Accessed April 15, 2024).
3. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the framingham study. *Stroke*. (1991) 22(8):983–8. doi: 10.1161/01.STR.22.8.983
4. López-López JA, Sterne JAC, Thom HHZ, Higgins JPT, Hingorani AD, Okoli GN, et al. Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. *Br Med J*. (2017) 359: j5058. doi: 10.1136/bmj.j5058
5. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J*. (2021) 42(5):373–498. doi: 10.1093/eurheartj/ehaa612
6. Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, et al. Atrial fibrillation in patients with cryptogenic stroke. *N Engl J Med*. (2014) 370(26):2467–77. doi: 10.1056/NEJMoa1311376
7. Lakkireddy DR, Garg J, Ahmed A, Bawa D, Kabra R, Darden D, et al. LB-456091-2 dynamic data driven management of atrial fibrillation using implantable cardiac monitors—the monitor af study. *Heart Rhythm*. (2023) 20(7):1085–6. doi: 10.1016/j.hrthm.2023.04.041
8. Hart RG, Diener H-C, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. (2014) 13(4):429–38. doi: 10.1016/S1474-4422(13)70310-7
9. Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. *Circulation*. (2024) 149(1):e1–e156. doi: 10.1161/CIR.0000000000001193
10. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nat Rev Cardiol*. (2021) 18(8):581–99. doi: 10.1038/s41569-021-00522-7
11. Dörr M, Nothmann V, Brasier N, Bosshard E, Djurdjevic A, Gross S, et al. The WATCH AF trial: smartWATCHes for detection of atrial fibrillation. *JACC Clin Electrophysiol*. (2019) 5(2):199–208. doi: 10.1016/j.jacep.2018.10.006
12. Chen W, Khurshid S, Singer DE, Atlas SJ, Ashburner JM, Ellinor PT, et al. Cost-effectiveness of screening for atrial fibrillation using wearable devices. *JAMA Health Forum*. (2022) 3(8):e222419. doi: 10.1001/jamahealthforum.2022.2419

Author contributions

ZP: Writing – original draft, Writing – review & editing. DS: Writing – original draft, Writing – review & editing. SM: Writing – original draft, Writing – review & editing. DA: Writing – original draft, Writing – review & editing. ME: Writing – original draft, Writing – review & editing. PK: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer VV declared a shared affiliation with the author PK to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

13. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol.* (2019) 12(6):e006834. doi: 10.1161/CIRCEP.118.006834
14. Tran KV, Filippaios A, Noorishirazi K, Ding E, Han D, Mohagheghian F, et al. False atrial fibrillation alerts from smartwatches are associated with decreased perceived physical well-being and confidence in chronic symptoms management. *Cardiol Cardiovasc Med.* (2023) 7(2):97–107. doi: 10.26502/fccm.92920314
15. Sepehri Shamloo A, Bollmann A, Dages N, Arya A, Hindricks G. Smart watch devices for atrial fibrillation screening. *J Am Coll Cardiol.* (2020) 75(11):1364–5. doi: 10.1016/j.jacc.2019.10.063
16. Charlton PH, Kyriaco PA, Mant J, Marozas V, Chowienicz P, Alastruey J. Wearable photoplethysmography for cardiovascular monitoring. *Proc IEEE Inst Electr Electron Eng.* (2022) 110(3):355–81. doi: 10.1109/JPROC.2022.3149785
17. Pereira T, Tran N, Gadhoumi K, Pelter MM, Do DH, Lee RJ, et al. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digit Med.* (2020) 3:3. doi: 10.1038/s41746-019-0207-9
18. Nilsson LM. Respiration signals from photoplethysmography. *Anesth Analg.* (2013) 117(4):859–65. doi: 10.1213/ANE.0b013e31828098b2
19. Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas.* (2007) 28(3):R1. doi: 10.1088/0967-3334/28/3/R01
20. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health.* (2017) 5:258. doi: 10.3389/fpubh.2017.00258
21. Foster KR, Torous J. The opportunity and obstacles for smartwatches and wearable sensors. *IEEE Pulse.* (2019) 10(1):22–5. doi: 10.1109/MPULS.2018.2885832
22. Bacevicius J, Abramikas Z, Dvinelis E, Audzijoniene D, Petrylaite M, Marinskienė J, et al. High specificity wearable device with photoplethysmography and six-lead electrocardiography for atrial fibrillation detection challenged by frequent premature contractions: doubleCheck-AF. *Front Cardiovasc Med.* (2022) 9:869730. doi: 10.3389/fcvm.2022.869730
23. Aschbacher K, Yilmaz D, Kerem Y, Crawford S, Benaron D, Liu J, et al. Atrial fibrillation detection from raw photoplethysmography waveforms: a deep learning application. *Heart Rhythm O2.* (2020) 1(1):3–9. doi: 10.1016/j.hroo.2020.02.002
24. Apple. Using Apple Watch for Arrhythmia Detection. (2020). Available online at: https://www.apple.com/healthcare/docs/site/Apples_Watch_Arrhythmia_Detection.pdf (Accessed April 15, 2024).
25. Lubitz SA, Faranesh AZ, Atlas SJ, McManus DD, Singer DE, Pagoto S, et al. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch: the fitbit heart study. *Am Heart J.* (2021) 238:16–26. doi: 10.1016/j.ahj.2021.04.003
26. University C. The Next Heartbeat. (2024). Available online at: <https://www.thenextheartbeat.com/definitions-2/> (Accessed April 15, 2024).
27. Mei Z, Gu X, Chen H, Chen W. Automatic atrial fibrillation detection based on heart rate variability and spectral features. *IEEE Access.* (2018) 6:53566–75. doi: 10.1109/ACCESS.2018.2871220
28. Hirsch G, Jensen SH, Poulsen ES, Puthusserypady S. Atrial fibrillation detection using heart rate variability and atrial activity: a hybrid approach. *Expert Syst Appl.* (2021) 169:114452. doi: 10.1016/j.eswa.2020.114452
29. Chen X, Cheng Z, Wang S, Lu G, Xv G, Liu Q, et al. Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ECG signals. *Comput Methods Programs Biomed.* (2021) 202:106009. doi: 10.1016/j.cmpb.2021.106009
30. Mousavi S, Afghah F, Acharya UR. HAN-ECG: an interpretable atrial fibrillation detection model using hierarchical attention networks. *Comput Biol Med.* (2020) 127:104057. doi: 10.1016/j.combiomed.2020.104057
31. Murat F, Sadak F, Yildirim O, Talo M, Murat E, Karabatak M, et al. Review of deep learning-based atrial fibrillation detection studies. *Int J Environ Res Public Health.* (2021) 18(21):11302. doi: 10.3390/ijerph182111302
32. Al-Kaisey AM, Koshy AN, Ha FJ, Spencer R, Toner L, Sajeev JK, et al. Accuracy of wrist-worn heart rate monitors for rate control assessment in atrial fibrillation. *Int J Cardiol.* (2020) 300:161–4. doi: 10.1016/j.ijcard.2019.11.120
33. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the apple watch. *J Med Syst.* (2019) 43(7):195. doi: 10.1007/s10916-019-1325-2
34. Koerber D, Khan S, Shamsheri T, Kirubakaran A, Mehta S. Accuracy of heart rate measurement with wrist-worn wearable devices in Various skin tones: a systematic review. *J Racial Ethn Health Disparities.* (2023) 10(6):2676–84. doi: 10.1007/s40615-022-01446-9
35. Kelly RI, Pearse R, Bull RH, Leveque JL, de Rigal J, Mortimer PS. The effects of aging on the cutaneous microvasculature. *J Am Acad Dermatol.* (1995) 33(5 Pt 1):749–56. doi: 10.1016/0190-9622(95)91812-4
36. Meijer WT, Hoes AW, Rutgers D, Bots ML, Hofman A, Grobbee DE. Peripheral arterial disease in the elderly: the rotterdam study. *Arterioscler Thromb Vasc Biol.* (1998) 18(2):185–92. doi: 10.1161/01.ATV.18.2.185
37. Niu Y, Wang H, Wang H, Zhang H, Jin Z, Guo Y. Diagnostic validation of smart wearable device embedded with single-lead electrocardiogram for arrhythmia detection. *Digit Health.* (2023) 9:20552076231198682. doi: 10.1177/20552076231198682
38. Jo YY, Cho Y, Lee SY, Kwon JM, Kim KH, Jeon KH, et al. Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *Int J Cardiol.* (2021) 328:104–10. doi: 10.1016/j.ijcard.2020.11.053
39. Steinberg JS, O'Connell H, Li S, Ziegler PD. Thirty-Second gold standard definition of atrial fibrillation and its relationship with subsequent arrhythmia patterns: analysis of a large prospective device database. *Circ Arrhythm Electrophysiol.* (2018) 11(7):e006274. doi: 10.1161/CIRCEP.118.006274
40. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med.* (2019) 381(20):1909–17. doi: 10.1056/NEJMoa1901183
41. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the fitbit heart study. *Circulation.* (2022) 146(19):1415–24. doi: 10.1161/CIRCULATIONAHA.122.060291
42. Avram R, Ramsis M, Cristal AD, Nathan V, Zhu L, Kim J, et al. Validation of an algorithm for continuous monitoring of atrial fibrillation using a consumer smartwatch. *Heart Rhythm.* (2021) 18(9):1482–90. doi: 10.1016/j.hrthm.2021.03.044
43. Badertscher P, Lischer M, Mannhart D, Knecht S, Isenegger C, Du Fay de Lavallaz J, et al. Clinical validation of a novel smartwatch for automated detection of atrial fibrillation. *Heart Rhythm O2.* (2022) 3(2):208–10. doi: 10.1016/j.hroo.2022.02.004
44. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, et al. Accuracy of apple watch for detection of atrial fibrillation. *Circulation.* (2020) 141(8):702–3. doi: 10.1161/CIRCULATIONAHA.119.044126
45. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. *Heart.* (2020) 106(9):665–70. doi: 10.1136/heartjnl-2019-316004
46. Mannhart D, Lischer M, Knecht S, du Fay de Lavallaz J, Strebel I, Serban T, et al. Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: bASEL wearable study. *JACC Clin Electrophysiol.* (2023) 9(2):232–42. doi: 10.1016/j.jacep.2022.09.011
47. Weidlich S, Mannhart D, Kennedy A, Doggart P, Serban T, Knecht S, et al. Reducing the burden of inconclusive smart device single-lead ECG tracings via a novel artificial intelligence algorithm. *Cardiovasc Digit Health J.* (2024) 5(1):29–35. doi: 10.1016/j.cvdhj.2023.12.003
48. Rizas KD, Freyer L, Sappler N, von Stülpnagel L, Spielbichler P, Krasniqi A, et al. Smartphone-based screening for atrial fibrillation: a pragmatic randomized clinical trial. *Nat Med.* (2022) 28(9):1823–30. doi: 10.1038/s41591-022-01979-w
49. McIntyre WF, Benz AP, Becher N, Healey JS, Granger CB, Rivard L, et al. Direct oral anticoagulants for stroke prevention in patients with device-detected atrial fibrillation: a study-level meta-analysis of the NOAH-AFNET 6 and ARTESIA trials. *Circulation.* (2024) 149(13):981–8. doi: 10.1161/CIRCULATIONAHA.123.067512
50. Kalla M, Fabritz L, Kirchhof P. SMART About watches: we need technical and biological validation of atrial fibrillation screening. *JACC Clin Electrophysiol.* (2019) 5(2):209–11. doi: 10.1016/j.jacep.2018.11.018
51. Kirchhof P, Camm AJ, Goette A, Brandes A, Eckardt L, Elvan A, et al. Early rhythm-control therapy in patients with atrial fibrillation. *N Engl J Med.* (2020) 383(14):1305–16. doi: 10.1056/NEJMoa2019422
52. Camm AJ, Naccarelli GV, Mittal S, Crijns HJGM, Hohnloser SH, Ma C-S, et al. The increasing role of rhythm control in patients with atrial fibrillation. *J Am Coll Cardiol.* (2022) 79(19):1932–48. doi: 10.1016/j.jacc.2022.03.337
53. Staerk L, Sherer JA, Ko D, Benjamin EJ, Helm RH. Atrial fibrillation: epidemiology, pathophysiology, and clinical outcomes. *Circ Res.* (2017) 120(9):1501–17. doi: 10.1161/CIRCRESAHA.117.309732
54. Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, et al. Screening for atrial fibrillation: uS preventive services task force recommendation statement. *JAMA.* (2022) 327(4):360–7. doi: 10.1001/jama.2021.23732
55. Hill NR, Groves L, Dickerson C, Ochs A, Pang D, Lawton S, et al. Identification of undiagnosed atrial fibrillation using a machine learning risk-prediction algorithm and diagnostic testing (PULSe-AI) in primary care: a multi-centre randomized controlled trial in England. *Eur Heart J Digit Health.* (2022) 3(2):195–204. doi: 10.1093/ehjdh/ztac009
56. Nadarajah R, Wu J, Hogg D, Raveendra K, Nakao YM, Nakao K, et al. Prediction of short-term atrial fibrillation risk using primary care electronic health records. *Heart.* (2023) 109(14):1072–9. doi: 10.1136/heartjnl-2022-322076
57. Noseworthy PA, Attia ZI, Behnken EM, Giblon RE, Bews KA, Liu S, et al. Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. *Lancet.* (2022) 400(10359):1206–12. doi: 10.1016/S0140-6736(22)01637-3
58. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. *J Am Coll Cardiol.* (2019) 74(19):2365–75. doi: 10.1016/j.jacc.2019.08.019

59. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. *N Engl J Med.* (2024) 390(12):1118–27. doi: 10.1056/NEJMra2307160
60. Wyatt KD, Poole LR, Mullan AF, Kopecky SL, Heaton HA. Clinical evaluation and diagnostic yield following evaluation of abnormal pulse detected using apple watch. *J Am Med Inform Assoc.* (2020) 27(9):1359–63. doi: 10.1093/jamia/ocaa137
61. Wang L, Nielsen K, Goldberg J, Brown JR, Rumsfeld JS, Steinberg BA, et al. Association of wearable device use with pulse rate and health care use in adults with atrial fibrillation. *JAMA Netw Open.* (2021) 4(5):e215821. doi: 10.1001/jamanetworkopen.2021.5821
62. Lee TT, Kesselheim AS. Food and drug administration precertification pilot program for digital health software: weighing the benefits and risks. *Ann Intern Med.* (2018) 168(10):730–2. doi: 10.7326/M17-2715
63. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. *JMIR Mhealth Uhealth.* (2019) 7(9):e12861. doi: 10.2196/12861
64. Services USDoHaH. The HIPAA Privacy Rule. (2022). Available online at: <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html> (Accessed April 15, 2024).
65. Gupta S, Mahmoud A, Massoomi MR. A clinician's guide to smartwatch "interrogation". *Curr Cardiol Rep.* (2022) 24(8):995–1009. doi: 10.1007/s11886-022-01718-0



OPEN ACCESS

EDITED BY

Evangelos Oikonomou,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Panteleimon Pantelidis,
National and Kapodistrian University of Athens,
Greece

Markus Zabel,
University Medical Center Göttingen, Germany

*CORRESPONDENCE

Rik Willems
✉ rik.willems@uzleuven.be

[†]These authors share senior authorship

RECEIVED 13 July 2024

ACCEPTED 03 October 2024

PUBLISHED 21 October 2024

CITATION

Ingelaere S, Villa A, Varon C, Van Huffel S,
Vandenberk B and Willems R (2024) AI-
enabled detection of QRS fragmentation from
12-lead electrocardiogram and its clinical
relevance for predicting malignant arrhythmia
onset.
Front. Cardiovasc. Med. 11:1464303.
doi: 10.3389/fcvm.2024.1464303

COPYRIGHT

© 2024 Ingelaere, Villa, Varon, Van Huffel,
Vandenberk and Willems. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

AI-enabled detection of QRS fragmentation from 12-lead electrocardiogram and its clinical relevance for predicting malignant arrhythmia onset

Sebastian Ingelaere^{1,2}, Amalia Villa³, Carolina Varon³,
Sabine Van Huffel³, Bert Vandenberk^{1,2†} and Rik Willems^{1,2*†}

¹Department of Cardiovascular Sciences, Faculty of Medicine, KU Leuven, Leuven, Belgium,

²Department of Cardiovascular Diseases, University Hospitals Leuven, Leuven, Belgium, ³Stadius Centre for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering, Faculty of Engineering Sciences, KU Leuven, Leuven, Belgium

Background: Electrocardiographic markers differentiating between death caused by ventricular arrhythmias and non-arrhythmic death could improve the selection of patients for implantable cardioverter-defibrillator (ICD) implantation. QRS fragmentation (fQRS) is a parameter of interest, but subject to debate. We investigated the association of an automatically quantified probability of fragmentation with the outcome in ICD patients.

Methods: From a single-center retrospective registry, all patients implanted with an ICD between January 1996 and December 2018 were eligible for inclusion. Patients with active pacing were excluded. From the electronic medical record, clinical characteristics at implantation were collected and a 12-lead ECG was exported and analyzed by a previously validated machine-learning algorithm to quantify the probability of fQRS. To compare fQRS(+) and fQRS(−) patients, dichotomization was performed using the Youden index. Patients with a high probability of fragmentation in any region (anterior, inferior or lateral), were labeled fQRS(+). The impact of this fQRS probability on outcomes was investigated using Cox regression.

Results: A total of 1,242 patients with a mean age of 62.6 ± 11.5 years and a reduced left ventricular ejection fraction of $31 \pm 12\%$ were included of which 227 (18.3%) were female. The vast majority suffered from ischemic heart disease (64.3%) and were implanted in primary prevention (63.8%). 538 (43.3%) had a high probability of fragmentation in any region. Patients with a high probability of fragmentation had more frequently dilated cardiomyopathy (39.4% vs. 33.0%, $p = 0.019$), left bundle branch block (40.8% vs. 32.5%, $p = 0.006$) and a higher use of cardiac resynchronization therapy with defibrillator (CRT-D) devices (33.9% vs. 26.3%, $p = 0.004$). After adjustment in a multivariable Cox model, there was no significant association between the probability of global or regional fQRS and appropriate ICD therapy, inappropriate shock and short- or long-term mortality.

Conclusion: There was no association between the automatically quantified probability of the presence of fQRS and outcome. This lack of predictive value might be due to the algorithm used, which identifies only the presence but not the severity of fragmentation.

KEYWORDS

QRS fragmentation, implantable-cardioverter defibrillator, sudden cardiac death, ventricular arrhythmias, artificial intelligence

Introduction

Cardiovascular mortality remains one of the leading causes of death worldwide. Fatal cardiac events can be classified by their underlying mechanism; either mechanical based on pump failure, or electrical secondary to ventricular arrhythmias. While drug therapy, with or without resynchronization pacing, is pivotal in preventing progression to pump failure, implantation of an implantable cardioverter-defibrillator (ICD) stands as a cornerstone in the prevention of sudden death due to ventricular arrhythmias (1). Despite the significantly increased risk of life-threatening ventricular arrhythmias in patients with heart failure, that are candidates for ICD implantation in primary prevention, the majority of patients with an ICD do not experience life-saving interventions from their device (2–4). In the current era of improved medical heart failure treatment and possible declining benefit of ICD therapy, optimal risk stratification is crucial to select patients at high-risk for ventricular arrhythmias but at low risk of non-arrhythmic death (5).

The electrocardiogram (ECG) is a widely available, easy-to-perform, low-cost and low-risk diagnostic tool essential in cardiology practice with potential to refine arrhythmia risk stratification. The QRS complex represents the ventricular depolarization on ECG. Myocardial scarring results in heterogeneous myocardial activation, causing conduction delay and creating substrates for reentry circuits. Fragmentation of the QRS complex was initially described by Flowers et al. in 1969 and implies a heterogeneous activation pattern (6). Despite its potential, consensus on the prognostic value of fragmented QRS (fQRS) remains elusive in literature, partly due to its typical binary use based on visual estimation, which is prone to significant inter- and intra-observer variability (7). We recently developed an algorithm to automatically quantify the probability of the presence of QRS fragmentation (8). In this study we aim to investigate the relationship between this probability of the presence of QRS fragmentation and outcomes of post-ICD implantation.

Methods

Patient selection

All patients with ischemic or dilated cardiomyopathy, who received their first ICD device at the University Hospitals of Leuven between 01/01/1996 and 31/12/2018 were eligible for inclusion (9). Both primary and secondary prevention ICD indications were included. Single and dual chamber ICD devices (VVI/DDD) as well as cardiac resynchronization defibrillator devices (CRT-D) were included. Patients with ventricular pacing on the ECG at time of implantation were excluded. Analysis on this retrospective dataset was approved by the local ethical committee of the University Hospitals of Leuven (S56074).

Quantification of QRS fragmentation

Pre-procedural resting 12-lead ECGs were recorded (paper speed of 25 mm/s, voltage of 10 mm/mV, sampling rate 250 Hz, filter range 0.05–150 Hz with notch filter at 50/60 Hz) and

analyzed using the “GE Marquette 12SL™ ECG Analysis Program” (GE Medical Systems, Menomonee Falls, WI, USA). Raw ECG data were exported for QRS fragmentation analysis. Recently, we reported on an AI-enabled supervised machine learning algorithm designed to quantify the probability of QRS fragmentation (8). This method employs a novel and robust QRS segmentation strategy combining information from multiple leads to ensure the inclusion of critical Q and S wave information while excluding artifactual oscillations and abnormal heartbeats. A dataset of pre-ICD implantation ECGs from 673 patients implanted at UZ Leuven was used (including both narrow and broad QRS complexes). Patients were on average 62.4 ± 11.6 years old and had a severely reduced left ventricular ejection fraction (LVEF) of $32.3 \pm 12.3\%$. The majority of patients were male (84.7%) and suffered from ischemic heart disease (66.3%). At the time of implantation, 57 patients (8.5%) were in atrial fibrillation, while the remaining ECGs exhibited sinus rhythm. These ECG recordings were evaluated lead-by-lead by five independent clinicians for presence or absence of QRS fragmentation using binary labeling. The sum of the scores per lead represented the level of agreement with signals scoring ‘0’ (full agreement that the signal was non-fragmented) and ‘5’ (full agreement that the signal was fragmented) used as the ground truth to train the model. Ten features as proposed by Goovaerts et al. were extracted from each signal (10). Three of these features were derived from phase-rectified signal averaging (PRSA) curves, which can detect oscillations in the QRS complex (mean derivative of the PRSA curve, slope and y-axis crossing of its linear fit). Six additional parameters were derived using variational mode decomposition (VMD), which decomposes signals into their frequency components. Lastly, the number of peaks within the QRS complex on the ECG was used. These ten features were used to train a Support Vector Machine (SVM), a machine learning classifier, to distinguish between fragmented and non-fragmented signals. Platt scaling was applied to convert the binary score into a continuous score by fitting a logistic regression, thus providing a probability of fragmentation. We used the UZ Leuven algorithm given the best performance in the UZ Leuven dataset (sensitivity 0.76, specificity 0.92, kappa 0.71). Output consisted of a continuous value between 0 and 1 for each signal, representing the probability of fragmentation. Mean values were calculated for anterior (V1–V5), lateral (I, aVL, V6) and inferior (II, III, aVF) regions.

Endpoints

Patients in the registry were followed every 6–12 months. Device related endpoints were 1-year, 3-year and overall appropriate therapy (i.e., anti-tachypacing and shock) or appropriate shock and ever inappropriate shock-rate. ICD interventions were analyzed by the treating cardiologist and registered in a standardized manner in the electronic medical record. Mortality related endpoints were 1-year, 3-year and overall mortality as well as ICD resistant mortality (ICD-RM). Guidelines recommend to reserve ICD therapy to patients with an expectation of good-quality survival of more than 1 year. Therefore, an ICD was considered clinically useful when

the ICD provided appropriate therapy and the patient survived implantation by 1 year and appropriate shock by 30 days. As such, to identify the subgroup of patients for whom ICD implantation proved futile, ICD-RM was defined as death within the first year after ICD implantation or within 30 days after a first shock or if death occurred without any documented appropriate ICD intervention during follow-up (4). Mortality was deducted from the national population register, linked to the electronic medical record, ensuring completeness of data. In case of heart transplantation, date of transplantation was considered last date of follow-up.

Statistical analysis

Categorical variables are presented as number with percentages and continuous variables as mean with standard deviation. Based on the receiver operating characteristic curve, cut-off for fragmentation was set at 0.642, the value with the highest Youden index (0.651) with at

least 95% specificity, as we want to include true fragmentation. Patients with or without fragmentation in any region were compared. We compared continuous variables by a Mann-Whitney *U* test and categorical variables by a Chi² test. *P*-values ≤ 0.05 were considered significant. Multicollinearity was tested by calculating the variance of inflation factors (VIFs). Univariate Cox regression analysis was performed to identify predictors of the predefined endpoints. For each endpoint, a multivariable Cox regression analysis was performed using the ENTER method with selection of variables based on a univariable *p*-value < 0.100. Statistical analysis was performed using SPSS Statistics (version 29, IBM Corp., Armonk, NY, USA).

Results

A total of 1,242 patients were included, with their baseline characteristics presented in Table 1. Almost half of them had QRS fragmentation in any region (*N* = 538, 43.3%). When comparing

TABLE 1 Baseline patient characteristics by QRS fragmentation.

QRS fragmentation <i>N</i> = 1,242	Overall		fQRS (–)		fQRS (+)		<i>p</i> -value
	Mean	SD	Mean	SD	Mean	SD	
Age (years)	62.6	11.5	62.9	10.6	61.1	12.9	0.165
BMI (kg/m ²)	26.4	4.4	26.6	4.4	26.6	4.3	0.856
LVEF (%)	31	12	31	11	30	11	0.425
Creatinine (mg/dl)	1.26	0.48	1.24	0.48	1.24	0.45	0.628
QTc (ms)	443	51	444	50	443	52	0.887
	Number	%	Number	%	Number	%	
Female	227	18.3	121	17.2	106	19.7	0.256
IHD (vs. DCM)	798	64.3	472	67.0	326	60.6	0.019
Secondary (vs. primary)	450	36.2	256	36.4	194	36.1	0.912
NYHA							
I	323	26.0	186	26.4	137	25.5	0.358
II	551	44.4	324	46.0	228	42.4	
III	359	28.9	190	27.0	169	31.4	
IV	8	0.6	4	0.6	4	0.7	
CRT-D (vs. VVI/DDD)	367	29.6	185	26.3	182	33.9	0.004
Stroke	134	10.8	72	10.2	62	11.5	0.465
DM	261	21.0	149	21.2	112	20.8	0.882
AF	355	28.6	192	27.3	163	30.4	0.234
AHT	665	53.5	390	55.4	275	51.1	0.134
BB	1,058	85.3	600	85.3	458	85.1	0.914
ACE-I or ARB	1,091	87.9	625	88.9	466	86.6	0.221
Loop	606	48.8	334	47.5	272	50.6	0.287
Aldactone	682	55.0	370	52.6	312	58.0	0.060
Anti-aggregant	748	60.3	439	62.4	309	57.4	0.074
Anticoagulation	420	33.8	226	32.1	194	36.1	0.144
Amiodarone	442	35.6	248	35.3	194	36.1	0.776
Sotalol	11	0.9	8	1.1	3	0.6	0.280
Digitalis	106	8.5	61	8.7	45	8.4	0.845
Statin	759	61.2	442	62.9	317	58.9	0.157
Ever appropriate	435	35.0	256	36.4	179	33.3	0.258
LBBB	386	36.2	192	32.5	194	40.8	0.006

Baseline patient characteristics with comparison between patients with [fQRS(+)] or without [fQRS(–)]. Age (years); body mass index, BMI (kg/m²); left ventricular ejection fraction, LVEF (%); creatinine (mg/dL); QTc interval, QTc (ms); ischemic heart disease, IHD; dilated cardiomyopathy, DCM; secondary vs. primary prevention indication; New York Heart Association class, NYHA; diabetes mellitus, DM; atrial fibrillation, AF; arterial hypertension, AHT; betablocker, BB; angiotensin converting enzyme inhibitor, ACE-I; angiotensin receptor blocker, ARB; loop diuretic, loop; ever appropriate ICD therapy (anti-tachypacing or shock), ever appropriate; left bundle branch block, LBBB; CRT-D, cardiac resynchronisation therapy with defibrillator; VVI/DDD, single and dual chamber ICD devices. Significant *p*-values in bold.

patients with or without QRS fragmentation at implantation, no significant disparities were observed. The average age of the cohort was 62.6 ± 11.5 years, without significant difference between patients with or without fragmentation. LVEF was comparable across patients with QRS fragmentation ($30 \pm 11\%$) as without fragmentation ($31 \pm 11\%$). Arterial hypertension ($N = 665$, 53.5%) and atrial fibrillation ($N = 355$, 28.6%) were the most common comorbidities with balanced prevalence between both groups. Only a minority of the included patients were female ($N = 227$, 18.3%). The sole noteworthy distinctions were a lower incidence of ischemic heart disease in patients with QRS fragmentation compared to those with dilated cardiomyopathy (60.6% vs. 67.0%, $p = 0.019$), a higher prevalence of CRT-D usage (33.9% vs. 26.3%, $p = 0.004$), and a greater frequency of LBBB (40.8% vs. 32.5%, $p = 0.006$).

QRS fragmentation and mortality

QRS fragmentation showed no predictive value for 1-year nor 3-year survival by univariate Cox regression analysis (Supplementary Appendix Tables S1–S2). Anterior region fragmentation appeared to predict ICD-RM (HR 1.668, $p = 0.036$) and approached significance for overall mortality (HR 1.467, $p = 0.058$) (Supplementary Appendix Tables S3–S4). Nonetheless, subsequent multivariable Cox modeling showed that anterior fragmentation lost its predictive power regarding ICD-RM (HR 1.112, $p = 0.730$). Significant predictors of ICD-RM were more pronounced heart failure symptoms, indicated by the detrimental impact of advanced NYHA status while there is a protective effect of CRT-D and better-preserved LVEF (Table 2). Cox-derived survival plot of ICD-RM stratified by anterior QRS fragmentation is presented in Figure 1. Overall mortality increased with age, higher creatinine, ischemic heart disease (IHD), worse NYHA status, most comorbidities and the use of loop diuretics and digitalis. A better preserved LVEF was associated with an improved overall survival (Table 3). Cox-derived survival plot of overall mortality stratified by anterior QRS fragmentation is presented in Figure 2.

QRS fragmentation and ICD therapy

In univariate Cox regression analysis, neither global nor regional QRS fragmentation exhibited any predictive value concerning clinically relevant device related endpoints, such as appropriate ICD intervention, appropriate ICD shock or inappropriate ICD shock (Supplementary Appendix Tables S5–S7). Secondary prevention and the use of digitalis were linked to an increased risk of appropriate ICD therapy, with or without shock (Supplementary Appendix Tables S8, S9). Conversely, patients treated with a statin exhibited a reduced risk of appropriate ICD shock (Supplementary Appendix Table S9). Even after adjustment, the presence of atrial fibrillation was associated with a higher likelihood of inappropriate ICD shocks (HR 1.992, $p < 0.001$). In contrast, older patients had a diminished risk of inappropriate ICD interventions (HR 0.981, $p = 0.010$) (Supplementary Appendix Table S10).

TABLE 2 Multivariable Cox regression regarding ICD resistant mortality (ICD-RM).

ICD-RM	HR	95% Confidence interval		p-value
		Lower	Upper	
Age (/year)	1.061	1.042	1.081	<0.001
BMI (/kg/m ²)	0.930	0.894	0.967	<0.001
LVEF (/%)	0.977	0.963	0.992	0.003
Creatinine (/mg/dl)	1.422	1.121	1.804	0.004
QTc (/ms)	1.001	0.998	1.003	0.673
Female	0.994	0.647	1.529	0.979
IHD (vs. DCM)	1.244	0.842	1.838	0.272
NYHA	ref			
I				
II	1.960	1.291	2.977	0.002
III	2.505	1.510	4.158	<0.001
IV	7.930	1.731	36.317	0.008
CRT-D (vs. VVI/DDD)	0.554	0.363	0.845	0.006
Stroke	1.274	0.807	2.009	0.298
DM	1.977	1.410	2.772	<0.001
AF	1.995	1.386	2.873	<0.001
AHT	0.930	0.681	1.269	0.646
Loop	1.535	1.094	2.153	0.013
Anticoagulation	0.836	0.580	1.207	0.340
Amiodarone	1.133	0.829	1.547	0.434
Statin	1.315	0.944	1.833	0.106
Anterior fQRS	1.112	0.609	2.029	0.730

Age (years); body mass index, BMI (kg/m²); left ventricular ejection fraction, LVEF (%); creatinine (mg/dL); QTc interval, QTc (ms); ischemic heart disease, IHD; dilated cardiomyopathy, DCM; New York Heart Association class, NYHA; diabetes mellitus, DM; atrial fibrillation, AF; arterial hypertension, AHT; loop diuretic, loop; CRT-D, cardiac resynchronisation therapy with defibrillator; VVI/DDD, single and dual chamber ICD devices.

Significant p -values in bold.

Discussion

The presence of QRS fragmentation has been extensively studied in various cardiac conditions as a potential marker of adverse outcomes (11–14). The current analysis found no association between the automatically quantified probability of QRS fragmentation and patient outcomes.

Prevalence of fQRS

The overall prevalence of fQRS in our cohort (43.3%) is higher compared to an analysis on MADIT II trial patients (33%) which might be due to differences in patient characteristics as the latter only included patients eligible for ICD in primary prevention (15). In a Finnish cohort of middle-aged subjects, the overall prevalence of fQRS was remarkably lower (19.7%). However, they included patients with and without cardiac antecedents (21.4% and 75.9% respectively) with a quasi-balanced cohort in terms of sex-category (52.3% men) and a notable lower average age (16). The same research group studied the prevalence of fQRS in patients without cardiac disease (19.0%) vs. ARTEMIS study patients with coronary artery disease with and without previous

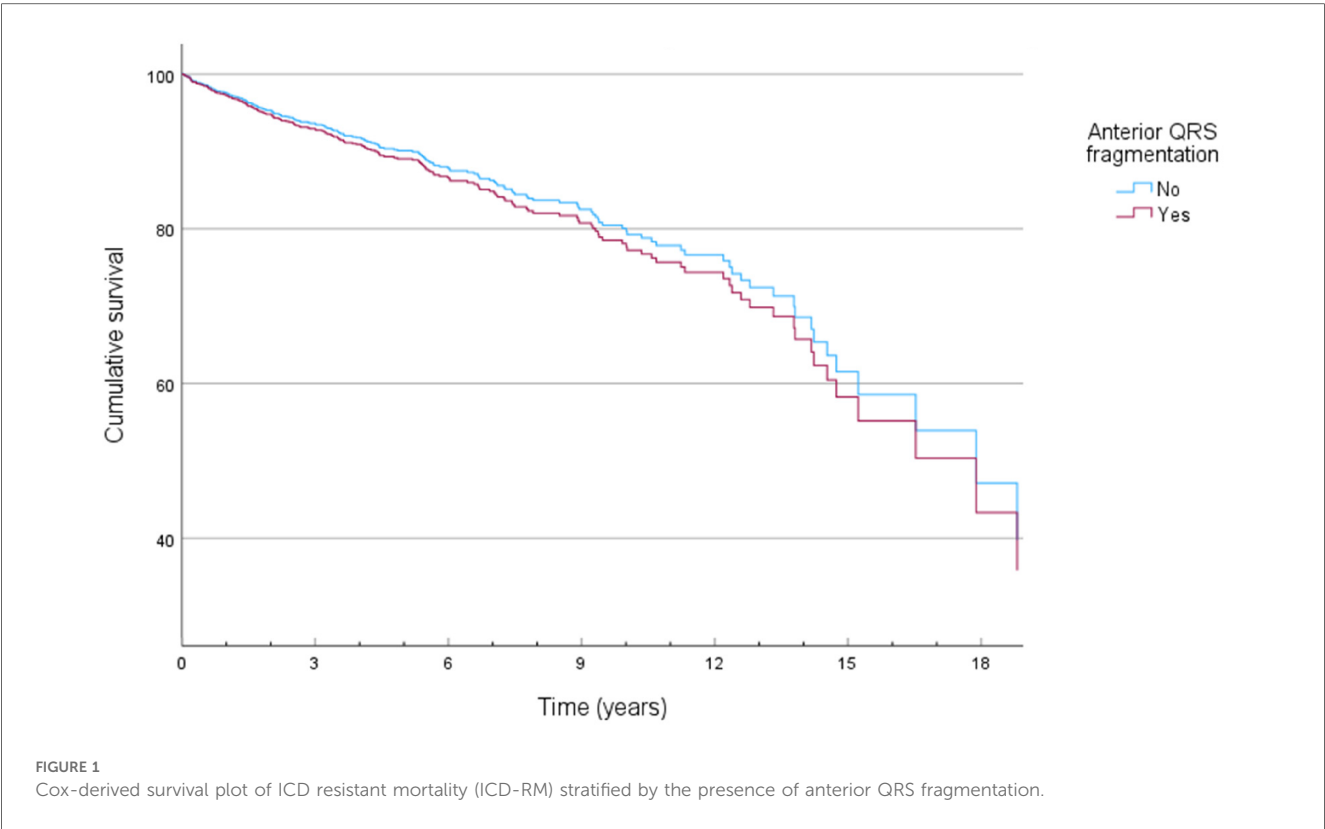


TABLE 3 Multivariable Cox regression regarding overall mortality.

Mortality (all)	HR	95% Confidence interval		p-value
		Lower	Upper	
Age (year)	1.050	1.036	1.063	<0.001
LVEF (/%)	0.979	0.968	0.989	<0.001
Creatinine (/mg/dl)	1.460	1.214	1.754	<0.001
QTc (/ms)	1.000	0.998	1.002	0.893
Female	0.939	0.680	1.295	0.699
IHD (vs. DCM)	1.536	1.173	2.012	0.002
Secondary (vs. primary)	1.240	0.966	1.590	0.091
NYHA	ref			
I				
II	1.602	1.220	2.103	0.001
III	1.461	1.054	2.027	0.023
IV	3.965	1.190	13.207	0.025
Stroke	1.400	1.007	1.947	0.045
DM	1.638	1.281	2.095	<0.001
AF	1.378	1.057	1.796	0.018
AHT	0.980	0.783	1.226	0.859
BB	0.833	0.640	1.084	0.174
Loop	1.396	1.096	1.778	0.007
Anticoagulation	0.931	0.713	1.215	0.598
Amiodarone	1.159	0.917	1.464	0.216
Digitalis	1.544	1.105	2.157	0.011
Anterior fQRS	1.034	0.655	1.631	0.887

Age (years); left ventricular ejection fraction, LVEF (%); creatinine (mg/dL); QTc interval, QTc (ms); ischemic heart disease, IHD; dilated cardiomyopathy, DCM; secondary vs. primary prevention indication; New York Heart Association class, NYHA; diabetes mellitus, DM; atrial fibrillation, AF; arterial hypertension, AHT; betablocker, BB; loop diuretic, loop. Significant p-values in bold.

myocardial infarction (39.5% and 35.3% respectively) vs. SCD victims from the FinGesture study (53.8%) (17). Accounting for the inclusion of secondary prevention patients (considered as aborted SCD victims) and the preponderance of IHD in our cohort, we assume the prevalence of fQRS in our study being representative. Inferior fragmentation was the most prevalent in our cohort as well as in all aforementioned study populations (15–17). In contrast to findings of Engstrom et al., we did not observe a significant difference in QRS fragmentation with age, nor was there a predisposition to fragmentation in males compared to females or any significant difference in the prevalence of diabetes or atrial fibrillation (14).

Predictive value of fQRS

In accordance with previous findings, there was no difference in NYHA functional status or the presence of ventricular arrhythmias, as defined by appropriate ICD therapy in patients with or without fQRS (18). The absence of significant predictive value of QRS fragmentation in this study contrasts with literature and diverges from our previous findings despite overlap in patients as this study is based on a single-center registry, updated until 2018 (19). Previous reports showed that the presence of fragmentation on ECG at hospital admission in acute STEMI patients was an independent predictor of in-hospital ventricular arrhythmias (20). Furthermore, a recent meta-analysis showed a significant prognostic value for fQRS in patients with acute

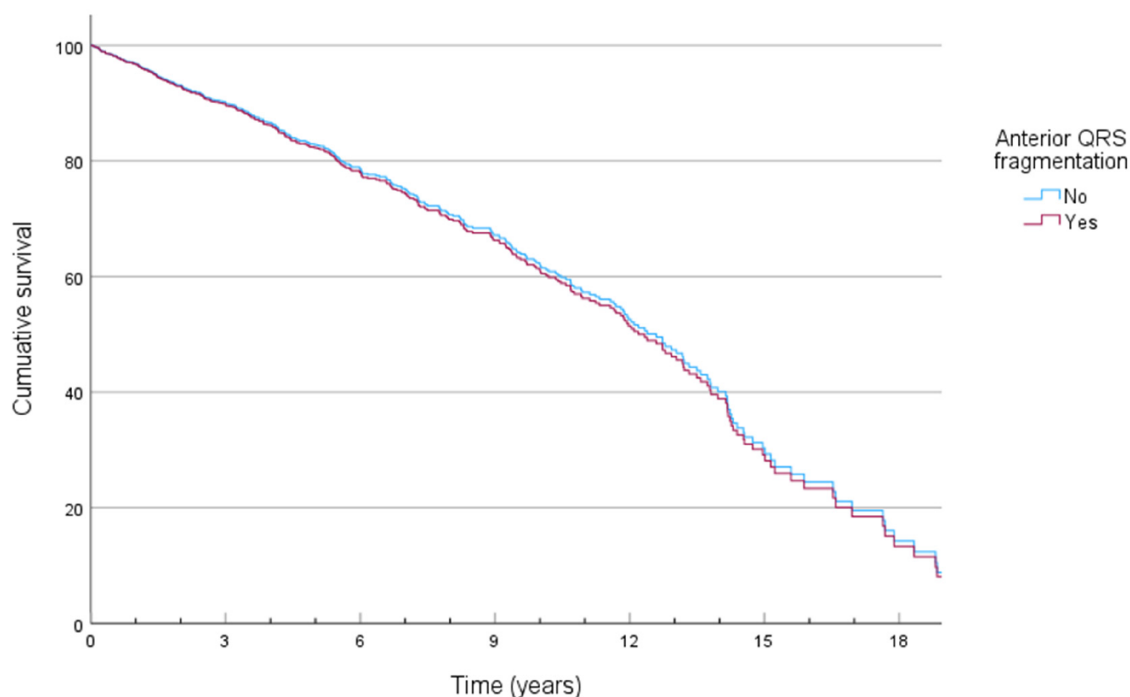


FIGURE 2
Cox-derived survival plot of overall mortality stratified by the presence of anterior QRS fragmentation.

myocardial infarction with an increased risk of in-hospital mortality and MACE, as well as long-term mortality and MACE (21). Another meta-analysis in heart failure patients also withheld fQRS as a predictor of ventricular arrhythmias but only when primary and secondary prevention patients were pooled as there was no significance in primary prevention alone (14). As patients with primary prevention ICD indications are most relevant for optimizing risk stratification, we performed a subgroup analysis but the results remained unchanged (Supplementary Appendix Table S11).

Further, our analysis did not establish an association between global nor regional QRS fragmentation and mortality. Terho et al. found an association between lateral QRS fragmentation and adverse survival rates in subjects with pre-existing cardiac diseases, while there was no association with mortality in healthy middle-aged subjects (17). Our findings contribute to the controversy surrounding fQRS. A meta-analysis by Rosengarten et al. found a significant association between fQRS and both all-cause mortality and SCD (22). However, more than half of the included studies were not significant with an important heterogeneity ($I^2 = 94\%$) resulting in a relative risk with broad 95% confidence interval [1.71 (1.02–2.85), $p = 0.04$] (22). The effect was mainly driven by two retrospective studies and one prospective study, the latter involving patients with higher LVEF, fewer comorbidities and only a small proportion of ICD recipients (23–25).

Despite the absence of predictive value of anterior QRS fragmentation concerning ICD-RM as well as overall mortality

after correction for confounders, several clinical parameters can still predict survival after ICD implantation. As previous research of our group demonstrated that ICD-RM occurred in up to a quarter of all patients after 5 years, it is not surprising that there is a significant overlap between the predictors for ICD-RM and overall mortality (4). Aging is inherently associated with a higher risk of mortality, regardless of any ICD interventions. Parameters reflecting a generally worse clinical condition—such as lower LVEF, higher creatinine (indicating poorer renal function), higher NYHA functional class and comorbidities like diabetes and atrial fibrillation—logically predict both ICD-RM and overall mortality. It is of interest that the use of CRT-D was associated with lower ICD-RM, possibly because addressing pump failure, reduces heart failure-related mortality, creating a window of opportunity for ventricular arrhythmias to occur. Subsequently, these ventricular arrhythmias can be terminated by the ICD device. While cardiac scar in IHD is known to be an excellent substrate for the development of re-entry phenomena leading to malignant ventricular arrhythmias, which can be terminated by an ICD, it also contributes to pump dysfunction and heart failure-related mortality, which is not modulated by a conventional ICD.

To ensure the robustness of our analysis, we checked for multicollinearity by calculating VIFs and found no evidence of multicollinearity among the variables included in the multivariable analysis. Furthermore, by including all clinically relevant covariables, we avoided specification bias.

Annotation and quantification of fQRS

Our findings should be interpreted with the limitations of visually scored QRS fragmentation in mind. Previous work of our group highlighted the inter- and intra-observer variability in interpretation of QRS fragmentation (7), prompting this current study aimed to assess the predictive value of QRS fragmentation in patients implanted with an ICD using an automated machine-learning algorithm (8). As our algorithm was trained with signals that were annotated by clinicians, it encompasses the aforementioned inter- and intra-observer variability. Furthermore, our model is trained with only ten preselected features, thus excluding other potential predictive features of the ECG signals. Therefore, in an effort to detect and quantify QRS fragmentation in a standardized and automated manner, the proposed algorithm actually estimates the likelihood that a clinician would perceive an ECG signal as fragmented based on 5 independent observers applying the Das criteria (8, 23, 26). Note that the original criteria proposed by Das et al. only distinguish between the presence or absence of fragmentation without assigning a numerical score reflecting the severity of fragmentation (26). To date, actual quantification of QRS fragmentation remains a challenge. Roudijk et al. attempted to quantify QRS fragmentation by scoring each deflection in the QRS complexes of 12 leads and found a significant association between higher scores and the presence of arrhythmogenic right ventricular cardiomyopathy, without being able to predict ventricular arrhythmias or SCD (27). In general, the results of visual QRS fragmentation should be interpreted with caution as the traditional visual ECG interpretation is inherently limited by subjectivity and observer variability emphasizing the importance of clear criteria to distinguish fragmentation due to myocardial scar from benign normal variants (7, 28). Therefore, computational quantification is increasingly pursued. Iglesias et al. approached this challenge using continuous wavelet transformation focusing on the frequency-domain instead of time-domain analysis (29). Another automated ECG-based evaluation of myocardial scar, the Selvester score, was significantly associated with mortality and device-related endpoints (30). However, the question arises whether all of these approaches identify fragmentation as described by Das or enclose other information. The lack of a uniform methodological approach compromises the reproducibility of studies in this field. Additionally, using data from 12 leads separately increases the risk of statistical errors. The latter advocates for dimensionality reduction, which reduces the risk of type I error by multiple testing and has the potential to enhance the signal-to-noise ratio. Recently, QRS micro-fragmentation was presented as a promising predictor of mortality (31). This analysis overcomes these hurdles as the analysis uses singular value decomposition of the QRS complex to project the information of all QRS complexes into optimized three perpendicular dimensions. Subsequently, the ECG is reconstructed back from this three-dimensional projection and micro-fragmentation is quantified as the

difference between the original and reconstructed ECG signals. By quantifying micro-fragmentation as percentage, it presents as a single, weighted value no longer prone to multicollinearity between different leads.

Limitations

All patients in this study were implanted in a tertiary hospital, which may introduce referral bias. Due to the design, general limitations inherent to retrospective, non-randomized research performed in a single center are valid. We only assessed the ECG at implantation not taking into account possible dynamical changes over time. Although we have reliable mortality data, we are blinded to the underlying mechanism of death while differentiation between arrhythmic and non-arrhythmic death might provide useful insights. Future work should focus on the true quantification of QRS fragmentation. As an extension of this research, improved prediction of outcomes might be possible by training new machine-learning models linking the current features directly to the outcome parameters, without the intermediate step of fQRS detection.

Conclusion

There was no association between the automatically quantified probability of QRS fragmentation and outcome when correcting for confounders by Cox proportional hazards modeling. While the algorithm quantifies the probability of QRS fragmentation, it does not directly assess the risk of arrhythmia or other adverse events. Instead, it uses fragmentation as an intermediate index, which might compromise its predictive value. Further research is warranted to improve quantification techniques and to establish a more robust understanding of the clinical relevance of fQRS in ICD patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the ethical committee of the University Hospitals of Leuven (S56074). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

SI: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. AV: Writing – review & editing, Methodology. CV: Writing – review & editing, Methodology. SH: Writing – review & editing, Methodology. BV: Writing – review & editing, Supervision, Methodology, Conceptualization. RW: Writing – review & editing, Supervision, Methodology, Conceptualization.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

RW reports research funding from Abbott, Biotronik, Boston Scientific, Medtronic; speakers and consultancy fees from Medtronic, Boston Scientific, Biotronik, Abbott. RW is supported as postdoctoral clinical researcher by the Fund for Scientific

Research Flanders. SH research received funding from the Flemish Government (AI Research Program). SH is affiliated to Leuven.AI—KU Leuven institute for AI, B-3000 Leuven, Belgium.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2024.1464303/full#supplementary-material>

References

- Zeppenfeld K, Tfelt-Hansen J, de Riva M, Winkel BG, Behr ER, Blom NA, et al. 2022 ESC guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) endorsed by the association for European paediatric and congenital cardiology (AEPC). *Eur Heart J*. (2022) 43(40):3997–4126. doi: 10.1093/eurheartj/ehac262
- Bigger J Jr. Why patients with congestive heart failure die: arrhythmias and sudden cardiac death. *Circulation*. (1987) 75(5 Pt 2):IV28–35.
- Koller MT, Schaer B, Wolbers M, Sticherling C, Bucher HC, Osswald S. Death without prior appropriate implantable cardioverter-defibrillator therapy: a competing risk study. *Circulation*. (2008) 117(15):1918–26. doi: 10.1161/CIRCULATIONAHA.107.742155
- Flore V, Vandenberk B, Belmans A, Garweg C, Ector J, Willems R. Introducing ICD-resistant mortality as an end point to evaluate the clinical efficacy of an implantable cardioverter-defibrillator in ischaemic cardiomyopathy. *Acta Cardiol*. (2018) 73(1):19–27. doi: 10.1080/00015385.2017.1322776
- Disertori M, Masè M, Rigoni M, Nollo G, Ravelli F. Declining clinical benefit of ICD in heart failure patients: temporal trend of mortality outcomes from randomized controlled trials. *J Cardiol*. (2020) 75(2):148–54. doi: 10.1016/j.jicc.2019.06.001
- Flowers NC, Horan LG, Tolleson WJ, Thomas JR. Localization of the site of myocardial scarring in man by high-frequency components. *Circulation*. (1969) 40(6):927–34. doi: 10.1161/01.CIR.40.6.927
- Vandenberk B, Robyns T, Goovaerts G, Claeys M, Helsen F, Van Soest S, et al. Inter- and intra-observer variability of visual fragmented QRS scoring in ischemic and non-ischemic cardiomyopathy. *J Electrocardiol*. (2018) 51(3):549–54. doi: 10.1016/j.jelectrocard.2017.12.002
- Villa A, Vandenberk B, Kenttä T, Ingelaere S, Huikuri HV, Zabel M, et al. A machine learning algorithm for electrocardiographic fQRS quantification validated on multi-center data. *Sci Rep*. (2022) 12(1):6783. doi: 10.1038/s41598-022-10452-0
- Vandenberk B, Garweg C, Voros G, Flore V, Marynissen T, Sticherling C, et al. Changes in implantation patterns and therapy rates of implantable cardioverter defibrillators over time in ischemic and dilated cardiomyopathy patients. *Pacing Clin Electrophysiol*. (2016) 39(8):848–57. doi: 10.1111/pace.12891
- Goovaerts G, Padhy S, Vandenberk B, Varon C, Willems R, Huffel SV. A machine-learning approach for detection and quantification of QRS fragmentation. *IEEE J Biomed Health Inform*. (2019) 23(5):1980–9. doi: 10.1109/JBHI.2018.2878492
- Das MK, Michael MA, Suradi H, Peng J, Sinha A, Shen C, et al. Usefulness of fragmented QRS on a 12-lead electrocardiogram in acute coronary syndrome for predicting mortality. *Am J Cardiol*. (2009) 104(12):1631–7. doi: 10.1016/j.amjcard.2009.07.046
- Pranata R, Yonas E, Vania R, Tondas AE, Yuniadi Y. Fragmented QRS is associated with intraventricular dyssynchrony and independently predicts nonresponse to cardiac resynchronization therapy-systematic review and meta-analysis. *Ann Noninvasive Electrocardiol*. (2020) 25(4):e12750. doi: 10.1111/anec.12750
- Basaran Y, Tigen K, Karaahmet T, Isiklar I, Cevik C, Gurel E, et al. Fragmented QRS complexes are associated with cardiac fibrosis and significant intraventricular systolic dyssynchrony in nonischemic dilated cardiomyopathy patients with a narrow QRS interval. *Echocardiography*. (2011) 28(1):62–8. doi: 10.1111/j.1540-8175.2010.01242.x
- Engstrom N, Dobson G, Ng K, Letson H. Fragmented QRS is associated with ventricular arrhythmias in heart failure patients: a systematic review and meta-analysis. *Ann Noninvasive Electrocardiol*. (2022) 27(1):e12910. doi: 10.1111/anec.12910
- Brenyo A, Pietrasik G, Barsheshet A, Huang DT, Polonsky B, McNitt S, et al. QRS Fragmentation and the risk of sudden cardiac death in MADIT II. *J Cardiovasc Electrophysiol*. (2012) 23(12):1343–8. doi: 10.1111/j.1540-8167.2012.02390.x
- Haukilahti MAE, Holmström L, Vähätalo J, Tikkanen JT, Terho HK, Kiviniemi AM, et al. Gender differences in prevalence and prognostic value of fragmented QRS complex. *J Electrocardiol*. (2020) 61:1–9. doi: 10.1016/j.jelectrocard.2020.05.010
- Terho HK, Tikkanen JT, Junttila JM, Anttonen O, Kenttä TV, Aro AL, et al. Prevalence and prognostic significance of fragmented QRS complex in middle-aged subjects with and without clinical or electrocardiographic evidence of cardiac disease. *Am J Cardiol*. (2014) 114(1):141–7. doi: 10.1016/j.amjcard.2014.03.066
- Lorgis L, Jourda F, Hachet O, Zeller M, Gudjoncik A, Dentan G, et al. Prognostic value of fragmented QRS on a 12-lead ECG in patients with acute myocardial infarction. *Heart Lung*. (2013) 42(5):326–31. doi: 10.1016/j.hrtlng.2013.05.005
- Vandenberk B, Robyns T, Goovaerts G, Van Soest S, Floré V, Garweg C, et al. Inferior and anterior QRS fragmentation have different prognostic value in patients who received an implantable defibrillator in primary prevention of sudden cardiac death. *Int J Cardiol*. (2017) 243:223–8. doi: 10.1016/j.ijcard.2017.02.131
- Attachaipanich T, Kittayaphong R. Fragmented QRS as a predictor of in-hospital life-threatening arrhythmic complications in ST-elevation myocardial

infarction patients. *Ann Noninvasive Electrocardiol.* (2019) 24(1):e12593. doi: 10.1111/anec.12593

21. Luo G, Li Q, Duan J, Peng Y, Zhang Z. The predictive value of fragmented QRS for cardiovascular events in acute myocardial infarction: a systematic review and meta-analysis. *Front Physiol.* (2020) 11:1027. doi: 10.3389/fphys.2020.01027

22. Rosengarten JA, Scott PA, Morgan JM. Fragmented QRS for the prediction of sudden cardiac death: a meta-analysis. *Europace.* (2015) 17(6):969–77. doi: 10.1093/europace/euu279

23. Das MK, Suradi H, Maskoun W, Michael MA, Shen C, Peng J, et al. Fragmented wide QRS on a 12-lead ECG: a sign of myocardial scar and poor prognosis. *Circ Arrhythm Electrophysiol.* (2008) 1(4):258–68. doi: 10.1161/CIRCEP.107.763284

24. Ozcan S, Cakmak HA, Ikitimur B, Yurtseven E, Stavileci B, Tufekcioglu EY, et al. The prognostic significance of narrow fragmented QRS on admission electrocardiogram in patients hospitalized for decompensated systolic heart failure. *Clin Cardiol.* (2013) 36(9):560–4. doi: 10.1002/clc.22158

25. Pei J, Li N, Gao Y, Wang Z, Li X, Zhang Y, et al. The J wave and fragmented QRS complexes in inferior leads associated with sudden cardiac death in patients with chronic heart failure. *Europace.* (2012) 14(8):1180–7. doi: 10.1093/europace/eur437

26. Das MK, Khan B, Jacob S, Kumar A, Mahenthiran J. Significance of a fragmented QRS complex versus a Q wave in patients with coronary artery disease. *Circulation.* (2006) 113(21):2495–501. doi: 10.1161/CIRCULATIONAHA.105.595892

27. Roudijk RW, Bosman LP, van der Heijden JF, de Bakker JMT, Hauer RNW, van Tintelen JP, et al. Quantitative approach to fragmented QRS in arrhythmogenic cardiomyopathy: from disease towards asymptomatic carriers of pathogenic variants. *J Clin Med.* (2020) 9(2):1–12. doi: 10.3390/jcm9020545

28. Haukilahti MA, Eranti A, Kenttä T, Huikuri HV. QRS Fragmentation patterns representing myocardial scar need to be separated from benign normal variants: hypotheses and proposal for morphology based classification. *Front Physiol.* (2016) 7:653. doi: 10.3389/fphys.2016.00653

29. García Iglesias D, Roqueñi Gutiérrez N, De Cos FJ, Calvo D. Analysis of the high-frequency content in human QRS complexes by the continuous wavelet transform: an automatized analysis for the prediction of sudden cardiac death. *Sensors (Basel).* (2018) 18(2):560. doi: 10.3390/s18020560

30. Reichlin T, Asatryan B, Vos MA, Willems R, Huikuri HV, Junttila MJ, et al. Automated electrocardiographic quantification of myocardial scar in patients undergoing primary prevention implantable cardioverter-defibrillator implantation: association with mortality and subsequent appropriate and inappropriate therapies. *Heart Rhythm.* (2020) 17(10):1664–71. doi: 10.1016/j.hrthm.2020.05.016

31. Hnatkova K, Andršová I, Novotný T, Britton A, Shipley M, Vandenberg B, et al. QRS micro-fragmentation as a mortality predictor. *Eur Heart J.* (2022) 43(40):4177–91. doi: 10.1093/eurheartj/ehac085



OPEN ACCESS

EDITED BY

Evangelos Oikonomou,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Emanuele Micaglio,
IRCCS San Donato Polyclinic, Italy
Giuseppe Mascia,
University of Genoa, Italy
Nicola Pierucci,
Sapienza University of Rome, Italy

*CORRESPONDENCE

Jian Zhou
✉ zj9758@126.com
Xiaorong Li
✉ jsxrl@yeah.net

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 08 August 2024

ACCEPTED 28 November 2024

PUBLISHED 16 December 2024

CITATION

Zhang B, Xie X, Yu J, Wu Y, Zhou J, Li X and Yang B (2024) A new prediction model for sustained ventricular tachycardia in arrhythmogenic cardiomyopathy. *Front. Cardiovasc. Med.* 11:1477931. doi: 10.3389/fcvm.2024.1477931

COPYRIGHT

© 2024 Zhang, Xie, Yu, Wu, Zhou, Li and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A new prediction model for sustained ventricular tachycardia in arrhythmogenic cardiomyopathy

Baowei Zhang^{1†}, Xin Xie^{1†}, Jinbo Yu¹, Yizhang Wu¹, Jian Zhou^{1*}, Xiaorong Li^{1,2*} and Bing Yang^{1,3}

¹Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, ²Department of Cardiology, Ji'an Center People's Hospital, Ji'an, China, ³Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Background: Arrhythmogenic cardiomyopathy (ACM) is an inherited cardiomyopathy characterized by high risks of sustained ventricular tachycardia (sVT) and sudden cardiac death. Identifying patients with high risk of sVT is crucial for the management of ACM.

Methods: A total of 147 ACM patients were retrospectively enrolled in the observational study and divided into training and validation groups. The least absolute shrinkage and selection operator (LASSO) regression model was employed to identify factors associated with sVT. Subsequently, a nomogram was constructed based on multivariable logistic regression analysis. The performance of the nomogram was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and calibration curve. Decision curve analysis was conducted to assess the clinical utility of the nomogram.

Results: Seven parameters were incorporated into the nomogram: age, male sex, syncope, heart failure, T wave inversion in precordial leads, left ventricular ejection fraction (LVEF), SDNN level. The AUC of the nomogram to predict the probability of sVT was 0.867 (95% CI, 0.797–0.938) in the training group and 0.815 (95% CI, 0.673–0.958) in the validation group. The calibration curve demonstrated a good consistency between the actual clinical results and the predicted outcomes. Decision curve analysis indicated that the nomogram had higher overall net benefits in predicting sVT.

Conclusion: We have developed and internally validated a new prediction model for sVT in ACM. This model could serve as a valuable tool to accurately identify ACM patients with high risk of sVT.

KEYWORDS

arrhythmogenic cardiomyopathy, sustained ventricular tachycardia, sudden cardiac death, prediction model, nomogram

Introduction

Arrhythmogenic cardiomyopathy (ACM), an inherited cardiomyopathy characterized by high risks of ventricular arrhythmias and sudden cardiac death, is one of the leading causes of SCD in young people and athletes (1–3). Sustained ventricular tachycardia (sVT) constitutes a major contributor to SCD in ACM patients. Although frequency catheter ablation (RFCA) and anti-arrhythmic drugs have effectively reduced sVT incidence, the implanted cardioverter defibrillator (ICD) remains the most efficacious

therapy for preventing SCD in ACM patients (4). However, ACM often presents in adolescents or at young age, and ICD treatment is invasive, associated with various complications, and imposes a dual burden of physiological and psychological stress on patients. Therefore, patients should be thoroughly assessed before undergoing ICD treatment (5). ACM patients with a history of sVT or resuscitated SCD can benefit from secondary prevention with ICD. Nevertheless, for ACM patients without documented sVT, the primary goal in the management of ACM patients is to identify those at high risk of sVT and treated with prophylactic ICD implantation (6).

Previous studies have identified several risk factors for predicting sVT in ACM patients, including gender, proband status, cardiac syncope, degree of exercise restriction, premature ventricular complex (PVC) burden, non-sustained VT, genetic testing, and extent of myocardial involvement (7, 8). Two studies have constructed prediction models for sudden cardiac death and sustained ventricular tachycardia in ACM patients using these risk factors (9, 10). These prediction models could effectively predict the occurrence of sVT and sudden cardiac death, but their predictive value was limited in ACM patients without identified pathogenic gene variants and in those with left ventricular involvement (11–13). These two prediction models considered demographic factors, symptoms, PVCs burden, and extent of myocardial involvement, but they overlooked the pathophysiological mechanisms underlying the occurrence of sVT in ACM patients. Increased cardiac sympathetic nerve activity plays an important role in the occurrence of sVT in ACM patients (14). Our preliminary study found that SDNN, an indicator of cardiac sympathetic nerve activity, could effectively predict the occurrence of sVT in ACM patients (15). The predictive value of SDNN is based on the role of hyperactivity of cardiac sympathetic nerve system in the occurrence of sVT, a mechanism distinct from other traditional risk factors. Therefore, we hypothesized that incorporating SDNN in the prediction model could provide additional value in distinguishing ACM patients with high risk of sVT.

In the present study, a total of 147 ACM patients were retrospectively enrolled and divided into training and validation groups. A prediction model was built on the basis of multivariable logistic regression analysis, and performed well in identifying ACM patients with high risk of sVT in both the training and validation groups.

Methods

Patients

A total of 212 patients diagnosed with ACM were retrospectively enrolled from the First Affiliated Hospital of Nanjing Medical University and Shanghai East Hospital between January 2006 and October 2024. This study was conducted in accordance with the principles outlined in the Declaration of Helsinki (revised in 2013) and received approval from the institutional ethics committee board of the First Affiliated

Hospital of Nanjing Medical University (Approval No. 2011-SR-014). Given the retrospective nature of the study, individual consent was waived. Diagnosis of ACM was based on the updated diagnostic criteria (6, 16). Patients with incomplete data, a history of heart transplantation (having received a normal heart), or other comorbidities associated with a high risk of ventricular arrhythmias (e.g., LVEF <35%) were excluded from the study.

Data collection

All enrolled ACM patients underwent retrospective evaluation of elementary clinical characteristics, as well as electrocardiographic and echocardiographic features. Elementary clinical data encompassed age, sex, age at diagnosis, prior syncope episodes, family history, comorbidities, and both pharmacological and non-pharmacological therapies. Surface electrocardiography (ECG), 24-h Holter monitoring, and echocardiography were meticulously reviewed by two independent physicians to identify potential predictors for risk stratification in ACM. The standard deviation of all normal-normal (NN) intervals (SDNN) was computed as the dispersion of all NN intervals over a 24-h period of Holter monitoring (17). Furthermore, all patients were meticulously assessed by two independent physicians to detect evidence of sVT. sVT was defined as follows: (1) recorded spontaneous persistent ventricular tachycardia (lasting for ≥ 30 s at ≥ 100 beats per minute, or with unstable hemodynamics requiring cardioversion), (2) ventricular fibrillation/flutter, or (3) appropriate ICD intervention in patients with such devices.

Statistical analysis

For continuous variables, the Kolmogorov-Smirnov test was employed to assess the distribution of the data. Variables exhibiting a normal distribution are expressed as mean $\pm$ standard deviation (SD), with differences between groups investigated using the Student's *t*-test. Variables not conforming to a normal distribution are presented with interquartile range and median, and differences were examined using the Mann-Whitney test. Categorical data were summarized using percentages, and differences between groups were evaluated using the χ^2 or Fisher's exact test, as appropriate.

The initial population was randomized into training and validation groups at a ratio of 7:3, which were subsequently utilized for model development and validation, respectively. The least absolute shrinkage and selection operator (LASSO) logistic regression model was employed to identify factors with significant predictive value for the occurrence of sVT from the data in the training group. Factors exhibiting nonzero coefficients in the LASSO regression model were retained. Subsequently, a multivariable logistic regression analysis, incorporating factors selected from the LASSO regression as well as other clinically significant variables, was conducted to construct a predictive model. To maximize the inclusion of effective predictive factors,

variables with a two-sided *p*-value ≤ 0.1 were incorporated into the prediction model, while factors of evident clinical significance were directly included. To provide quantitative predictions of sVT probability for individual patients, a nomogram was constructed based on the multivariable logistic regression analysis conducted in the training group.

To evaluate the discriminatory capability of the nomogram, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and calibration curve analyses were conducted in the training group. These methods were also applied to the validation group to further validate the performance of the nomogram. Additionally, decision curve analyses (DCA) were performed to assess the clinical utility of the nomogram by quantifying the net benefits in both groups. All statistical analyses were conducted using R software (Version 4.2.0; <https://www.r-project.org>).

Results

Baseline characteristics

After excluding 65 patients for various reasons, a total of 147 ACM patients were ultimately enrolled in this study (Supplementary Figure S1). Among them, 60.5% of patients (89 individuals) had a history of sVT. Patients with and without sVT exhibited similar characteristics in terms of age, history duration, age at diagnosis, family history of ACM, comorbidities, and pharmacological therapies. However, patients with sVT demonstrated a higher proportion of males (87.6% vs. 56.9%, $p < 0.001$), smokers (29.2% vs. 6.9%, $p = 0.002$), alcohol consumers (15.7% vs. 3.4%, $p = 0.04$), users of class III anti-arrhythmic drugs (49.4% vs. 19.0%, $p < 0.001$), recipients of ICD devices (34.8% vs. 10.3%, $p = 0.002$), individuals who underwent RFCA (34.8% vs. 12.1%, $p = 0.004$), those with T wave inversion in anterior leads (73.0% vs. 48.3%, $p = 0.004$), right ventricular (RV) enlargement (68.5% vs. 32.8%, $p < 0.001$), and lower levels of SDNN [113.0 (89.0, 135.0) vs. 134.5 (113.0, 150.5) ms, $p < 0.001$], compared with patients without sVT (Table 1).

All 147 ACM patients were subsequently randomly divided into the training group ($n = 104$, 70.7%), with the remaining patients allocated to the validation group ($n = 43$, 29.3%). The incidence of sVT was found to be similar between the training and validation groups (56.7% vs. 69.8%, $p = 0.20$). Moreover, demographic characteristics, comorbidities, pharmacological therapies, electrocardiographic, and echocardiographic features were well-balanced between the training and validation groups (Supplementary Table S1).

Feature selection and nomogram construction

In the training group, patients with sVT exhibited a higher proportion of males (93.2% vs. 55.6%, $p < 0.001$), smokers (32.2% vs. 4.4%, $p = 0.001$), alcohol consumers (18.6% vs. 0, $p = 0.006$), recipients of ICD devices (32.2% vs. 11.1%, $p = 0.02$), individuals

TABLE 1 Baseline characteristics of patients enrolled.

Variables	sVT group (<i>n</i> = 89, 60.5%)	Non-sVT group (<i>n</i> = 58, 39.5%)	<i>P</i> -value
Age (years)	45.9 ± 15.3	45.0 ± 15.8	0.72
Male (<i>n</i> , %)	78 (87.6)	33 (56.9)	<0.001
History (years)	3.0 (1.0, 10.0)	4.0 (1.0, 8.5)	0.91
Age of diagnosis (years)	42.0 (28.5, 48.0)	38.0 (28.5, 51.0)	0.93
Syncope (<i>n</i> , %)	39 (43.8)	21 (36.2)	0.46
ICD (<i>n</i> , %)	31 (34.8)	6 (10.3)	0.002
RFCA (<i>n</i> , %)	31 (34.8)	7 (12.1)	0.004
Family history (<i>n</i> , %)	8 (9.0)	3 (5.2)	0.59
Comorbidities			
Hypertension (<i>n</i> , %)	10 (11.2)	10 (17.2)	0.43
Diabetes (<i>n</i> , %)	0 (0)	3 (5.2)	0.12
Coronary artery disease (<i>n</i> , %)	2 (2.2)	0 (0)	0.67
Heart failure (<i>n</i> , %)	7 (7.9)	9 (15.5)	0.24
Smoking (<i>n</i> , %)	26 (29.2)	4 (6.9)	0.002
Alcohol (<i>n</i> , %)	14 (15.7)	2 (3.4)	0.04
Anti-arrhythmic drugs			
Class I (<i>n</i> , %)	15 (16.9)	7 (12.1)	0.58
Class II (<i>n</i> , %)	43 (48.3)	18 (31.0)	0.06
Class III (<i>n</i> , %)	44 (49.4)	11 (19.0)	<0.001
Class IV (<i>n</i> , %)	3 (3.4)	0 (0)	0.41
Characteristics on electrocardiography			
RBBB (<i>n</i> , %)	33 (37.1)	18 (31.0)	0.57
TWI-A (<i>n</i> , %)	65 (73.0)	28 (48.3)	0.004
TWI-I (<i>n</i> , %)	38 (42.7)	19 (32.8)	0.30
NSVT (<i>n</i> , %)	39 (43.8)	25 (43.1)	1.00
SDNN (ms)	113.0 (89.0, 135.0)	134.5 (113.0, 150.5)	<0.001
Characteristics on echocardiography			
RV enlargement (<i>n</i> , %)	61 (68.5)	19 (32.8)	<0.001
RVOT dyskinesia (<i>n</i> , %)	40 (44.9)	31 (53.4)	0.40
RV free wall dyskinesia (<i>n</i> , %)	63 (70.8)	34 (58.6)	0.18
PAH (<i>n</i> , %)	32 (36.0)	13 (22.4)	0.12
Tricuspid regurgitation (<i>n</i> , %)	63 (70.8)	32 (55.2)	0.08
Mitral regurgitation (<i>n</i> , %)	26 (29.2)	19 (32.8)	0.79
LVDD (mm)	47.0 (41.5, 50.0)	48.0 (45.0, 51.0)	0.09
LVEF (%)	61.0 (52.5, 65.8)	62.9 (55.0, 65.8)	0.59

sVT, sustained ventricular tachycardia; ICD, implantable cardiac defibrillator; RFCA, radiofrequency catheter ablation; RBBB, right bundle branch block; TWI-A, T wave inversion on anterior leads; TWI-I, T wave inversion on inferior leads; NSVT, non-sustained ventricular tachycardia; SDNN, Standard deviation of NN intervals; RV, right ventricle; PAH, pulmonary artery hypertension; LVDD, left ventricular end diastolic dimension; LVEF, left ventricular ejection fraction.

who underwent RFCA (39.0% vs. 8.9%, $p = 0.001$), users of class III anti-arrhythmic drugs (AAD) therapy (54.2% vs. 24.4%, $p = 0.004$), those with T wave inversion in anterior leads (74.6% vs. 51.1%, $p = 0.02$), and RV enlargement (71.2% vs. 37.8%, $p = 0.001$), but exhibited lower levels of SDNN [114.0 (88.0, 136.0) vs. 132.0 (115.0, 151.0) ms, $p = 0.003$], consistent with findings observed in the entire cohort (Supplementary Table S2). The LASSO logistic regression model identified 11 features as potential predictors of sVT in the training group (Figure 1). Subsequently, a multivariable logistic regression

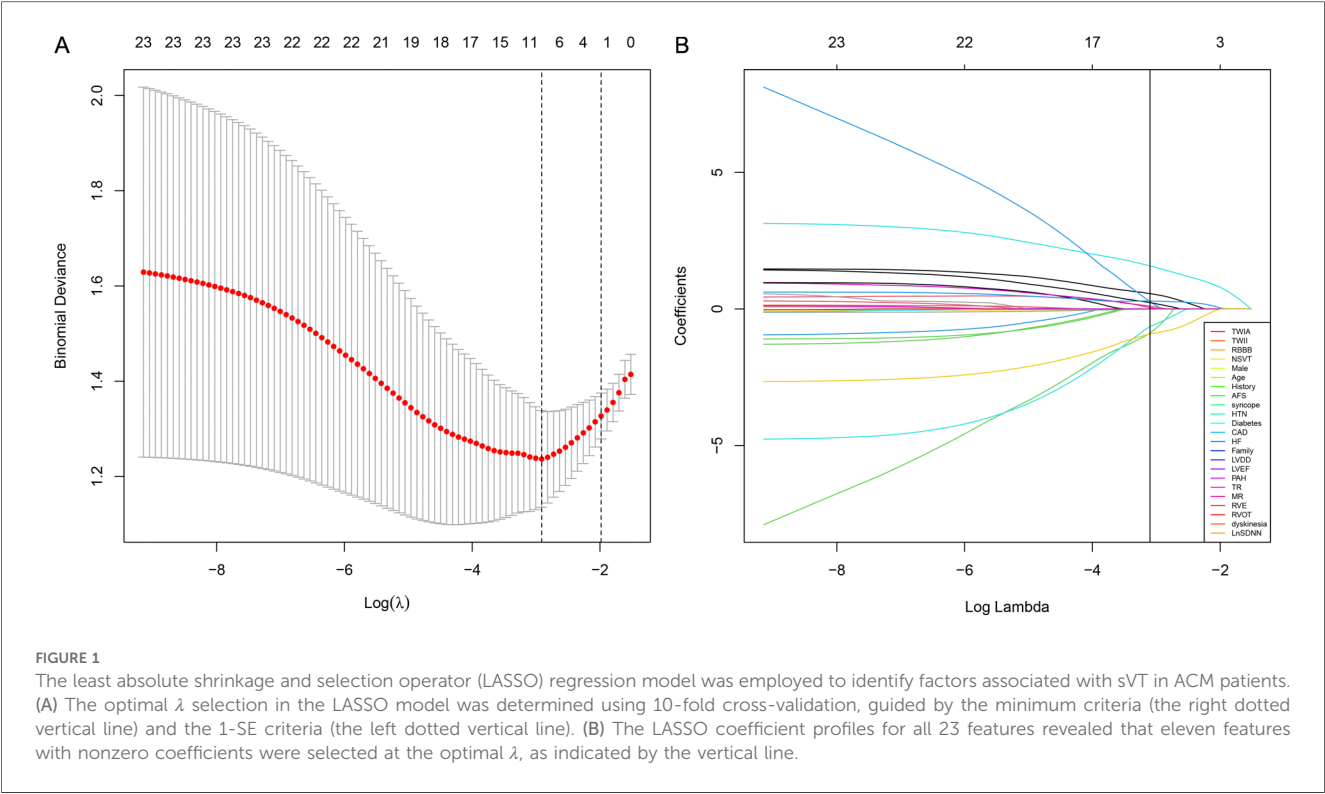


TABLE 2 Predictors for sVT in patients with ACM.

Variable	Prediction model		
	β	Odds ratio (95% CI)	P-value
TWI-A	1.38	3.97 (1.148–15.647)	0.04
LnSDNN	–2.19	0.11 (0.011–0.813)	0.04
Male	2.81	16.62 (3.613–108.370)	<0.001
Heart failure	–4.43	0.01 (0.001–0.155)	0.002
LVEF	–0.11	0.90 (0.818–0.978)	0.02
Syncope	1.01	2.73 (0.698–11.527)	0.15
Age	0.02	1.02 (0.970–1.071)	0.44

sVT, sustained ventricular tachycardia; ACM, arrhythmogenic cardiomyopathy; CI, confidence interval; TWI-A, T wave inversion on anterior leads; TWI-I, T wave inversion on inferior leads; NSVT, non-sustained ventricular tachycardia; LVDD, left ventricular end diastolic dimension; LVEF, left ventricular ejection fraction; RV, right ventricle; PAH, pulmonary artery hypertension; NSVT, non-sustained ventricular tachycardia; SDNN, Standard deviation of NN intervals; NA, not available.

analysis based on the factors selected in the LASSO regression was conducted to construct the prediction model. Additionally, for the comprehensive predictive value of sVT in ACM patients, syncope and left ventricular ejection fraction (LVEF) were also included. Ultimately, age, syncope, LVEF, male sex, heart failure, T wave inversion in anterior leads, and SDNN were incorporated into the prediction model (Table 2).

Based on the predictors identified by the LASSO logistic regression model in the training group, a nomogram was constructed in the training group for identifying ACM patients at high risk of sustained ventricular tachycardia (sVT). The nomogram included five significant predictive features and two canonical predictors (Figure 2). Each factor in the nomogram

was assigned an individual score based on its value, and a total score was calculated by summing the scores of all factors. The final score derived from the nomogram could be utilized to estimate the risk of sVT for a given patient (Figure 2).

Performance of the nomogram

We firstly validated the predictive value of the nomogram for identifying ACM patients at high risk of sVT in the training group by means of the ROC curve. The AUC of the nomogram for predicting the occurrence of sVT in ACM patients was 0.867 (95% CI, 0.797–0.938) (Figure 3A). The calibration curve demonstrated that the predictive nomogram for the occurrence of sVT provided excellent estimations of actual probabilities (Figure 3B).

Subsequently, the performance of the prediction model was validated in the validation group. Among the 43 patients allocated to the validation group, 30 had a history of sVT (69.8%). Patients with sVT in the validation exhibited similar proportions of males, smokers, recipients of ICD devices, individuals who underwent RFCA, and those with T-wave inversion in anterior leads, which was different from patients in the training group (Supplementary Table S3). However, the AUC of the prediction model to forecast the incidence of sVT was 0.815 (95% CI, 0.673–0.958) in the validation group (Figure 3C), indicating that the prediction model demonstrated excellent performance in identifying ACM patients at high risk of sVT. Furthermore, the calibration curve also illustrated good consistency between the actual incidence of sVT and the predicted probability in the validation group (Figure 3D).

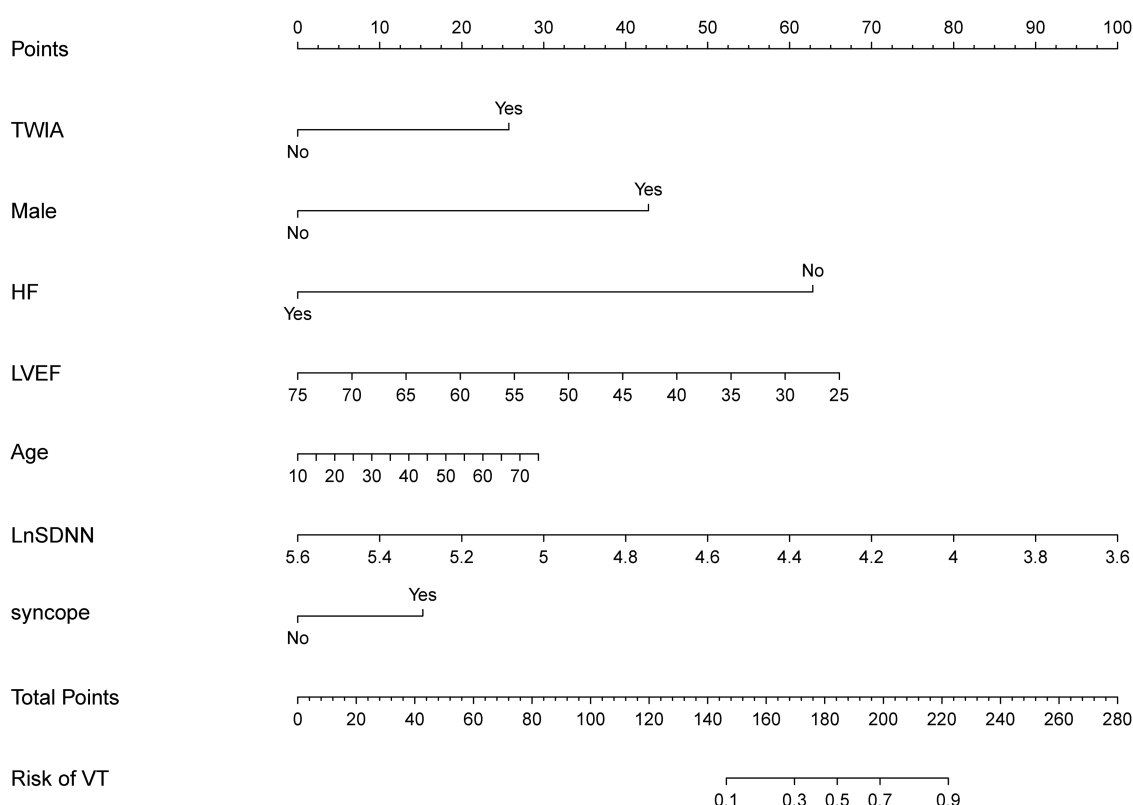


FIGURE 2

The nomogram was constructed in the training group for identifying ACM patients at high risk of sVT. Each factor in the nomogram was assigned an individual score based on its value, and the scores of all factors were summarized to a total score, which was used to estimate the risk of sVT for a given patient.

Clinical utility

To explore the clinical implications of our predictive model, decision curve analyses were conducted in both the training group and the validation group. As depicted in Figure 4, the results of DCA illustrated that the predictive model provided excellent overall net benefits in predicting the incidence of sVT in ACM patients.

Discussion

In this study, we developed and internally validated a prediction model for identifying ACM patients at high risk of sVT. All features incorporated into the prediction model consisted of elementary clinical information readily available in clinical practice, thus rendering the prediction model suitable for routine use in clinical settings. Predicted risks were consistent with the actual risks of sVT in both the training group [AUC 0.867 (95% CI, 0.797–0.938)] and the validation group [AUC 0.815 (95% CI, 0.673–0.958)]. Furthermore, decision curve analysis also demonstrated that the model was associated with a high net benefit in predicting the incidence of sVT in ACM patients.

High risk of sVT is the prominent clinical feature in patients with ACM. Previous studies have shown that patients with ACM

exhibit annual sVT rates ranging from 2% to 10%, which are notably higher compared to patients with other non-ischemic cardiomyopathies (18–20). Substitution of the ventricular myocardium with fibrous or fibrofatty tissue represents a distinctive pathological feature in patients with ACM (21). The infiltration of fibrous or fibrofatty tissue disrupts the conduction velocity and direction of cardiac electrical activity, predisposing individuals to the formation of reentrant circuits and thereby increasing the risk of sVT (22). Premature ventricular beats and sVT exhibiting a left bundle branch block (LBBB) morphology with a non-inferior axis pattern in ACM patients indicate the “triangle of dysplasia” as the origin of ventricular arrhythmias, thereby supporting cardiac fibrosis as an important mechanism of sVT in ACM (23). However, clinical studies have shown that VAs could manifest before identifiable cardiac structural changes occur, suggesting the involvement of other mechanisms contributing to ventricular arrhythmias in ACM. Subsequent studies identified impairment of action potential upstroke velocity and propagation of action potential in cardiomyocytes from ACM mice, attributed to the dislocation of connexin 43 (Cx43) and dysfunction of sodium channel Nav1.5 (24–26). Furthermore, dysregulation of calcium (Ca^{2+}) handling has been observed in cardiomyocytes from ACM mice (27). This disruption of Ca^{2+} homeostasis predisposes to sarcoplasmic reticulum spontaneous Ca^{2+} release and early afterdepolarizations, ultimately leading to increased cardiomyocyte

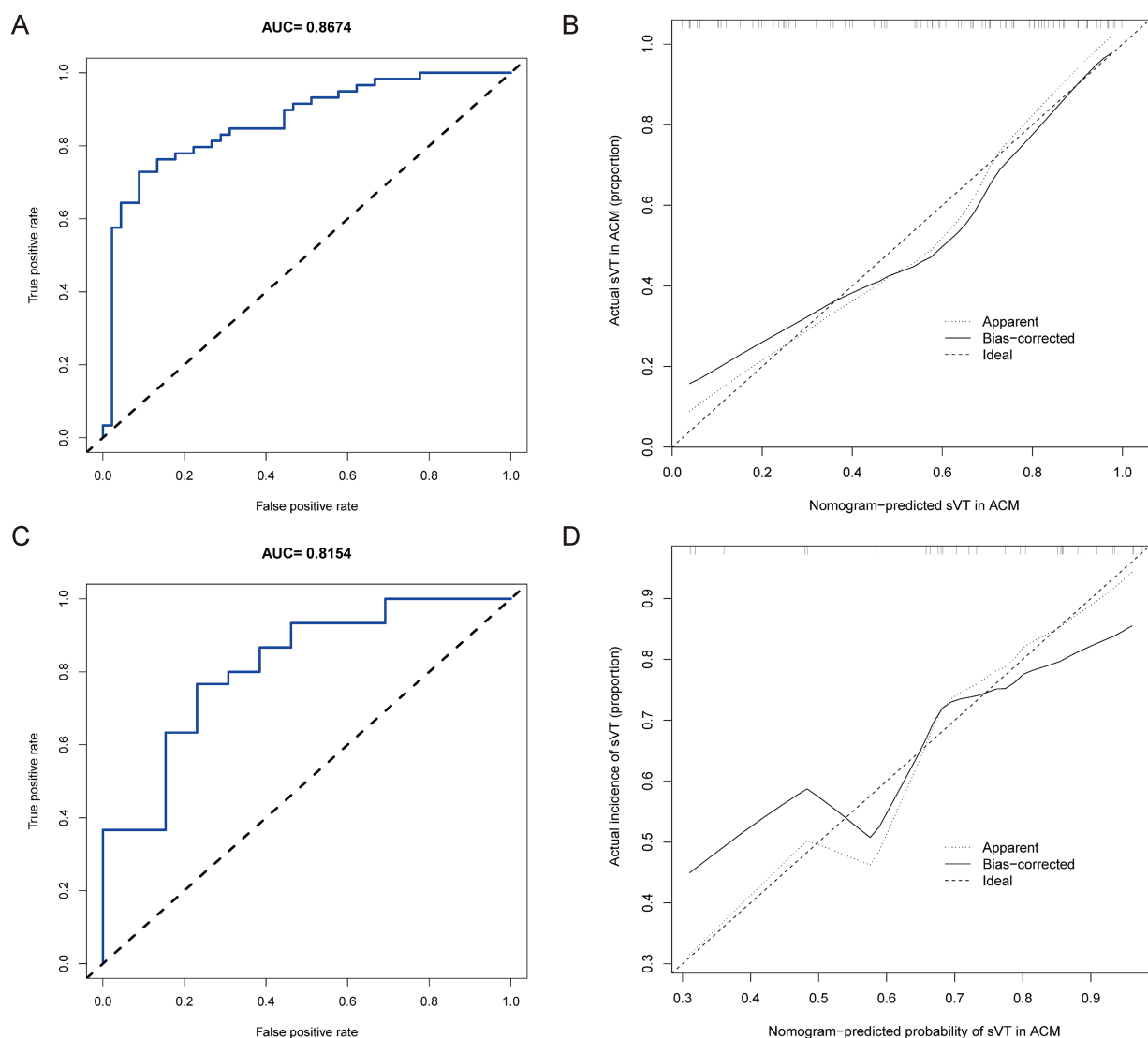


FIGURE 3

The nomogram was validated in both the training and validation groups. (A) The ROC curve was utilized to evaluate the performance of the nomogram in the training group. The AUC of the nomogram for predicting the occurrence of sVT in ACM patients was 0.867 (95% CI, 0.797–0.938). (B) The calibration curve in the training group demonstrated that the predictive nomogram provided excellent estimations of actual probabilities of sVT. (C) The ROC curve was used to evaluate the performance of the nomogram in the validation group. The AUC of the nomogram for predicting the occurrence of sVT in ACM patients was 0.815 (95% CI, 0.673–0.958). (D) The calibration curve in the validation group validated that the predictive nomogram provided excellent estimations of actual probabilities of sVT.

excitability and the onset of ventricular arrhythmogenesis (28). Consequently, a combination of cardiac fibrosis, dislocation of Cx43 and Nav1.5, and disturbance of Ca^{2+} homeostasis is postulated to contribute to arrhythmogenesis in ACM.

In addition, the abnormal activity of sympathetic nerve system contributes to the ventricular arrhythmogenesis in ACM (29). Our clinical studies, along with others, have identified the overactivation of the cardiac sympathetic system in patients with ACM (14, 15, 30). Stimulation of myocardial β_1 adrenergic receptors can increase cellular Ca^{2+} loading through the cyclic adenosine monophosphate (cAMP)–protein kinase A (PKA) pathway, resulting in an augmentation in the L-type calcium current, as well as through SERCA2-mediated reduction of Ca^{2+} re-uptake. The resultant

elevation in diastolic Ca^{2+} levels within the cardiomyocyte may heighten the likelihood of spontaneous sarcoplasmic reticulum calcium release, subsequently leading to the induction of afterdepolarization and extra stimulus, which could serve as triggers for inducing sVT (31). Additionally, the sprouting and remodeling of sympathetic nerve fibers secondary to sympathetic nerve fiber necrosis result in the heterogeneous reinnervation of cardiomyocytes, potentially promoting the dispersion of action potential duration and providing a suitable substrate for the maintenance of sVT (32). Therefore, the overactivity of the cardiac sympathetic nervous system not only provides triggers but also a suitable substrate for sVT in ACM patients.

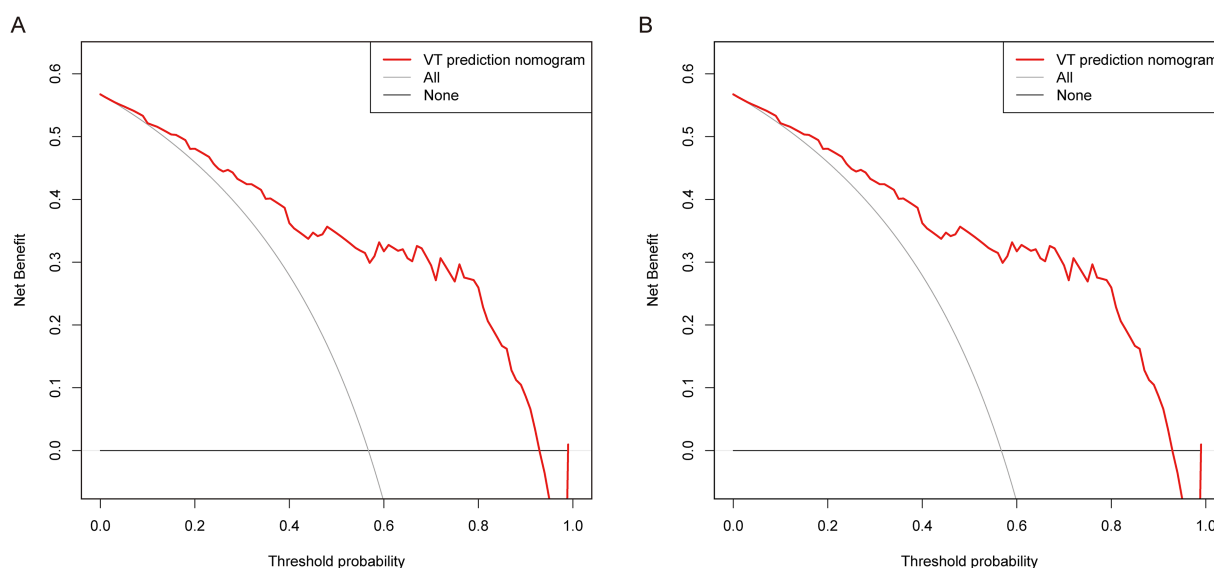


FIGURE 4

DCA was performed for the nomogram in both the training group (A) and the validation group (B) the red line indicated the predictive nomogram. The gray line represented the assumption that sVT occurred in all ACM patients, while the black line represented the assumption that no patient had sVT.

Sustained ventricular tachycardia in ACM patients is associated with abnormal electroactivity of cardiomyocytes, cardiac fibrosis, and overactivation of the cardiac sympathetic system. Therefore, clinical parameters associated with these factors may serve as predictors of sVT in ACM patients. Previous studies have identified several predictors of sVT in ACM patients, including recent syncope, younger age at presentation, male sex, decreased left ventricular ejection fraction (LVEF), non-sustained ventricular tachycardia (NSVT), PVC burden, RV dysfunction, and leads with T-wave inversion (33–36). Syncope is a common initial symptom of ACM and often indicates that the patient has experienced hemodynamically unstable sVT, although non-cardiogenic causes of syncope should also be considered in some cases (37). Therefore, recent syncope is regarded as a significant indicator for identifying high-risk patients with sVT in ACM. NSVT and PVC burden may reflect the abnormal electroactivity of cardiomyocytes. Decreased LVEF and RV dysfunction may indicate more severe cardiac fibrosis and scar formation in both ventricles (38). Additionally, the 12-lead ECG plays a crucial role in the diagnosis of ACM, localization of ventricular arrhythmias, and identification of high-risk patients for sVT (39). T-wave inversion is not only one of the ECG criteria for diagnosing ACM, but it is also an independent factor in predicting the occurrence of sVT in patients with ACM (20, 40). This study also identified that T-wave inversion in the anterior leads was independently associated with sVT in ACM patients. A recent study integrated these predictors and developed a prediction model for sVT in ACM patients (10). However, subsequent study showed that this prediction model had good performance in distinguishing patients with and without sVT in ACM patients with pathogenic mutant *PKP2* (C-index 0.77), but had limited value in predicting the risk of sVT in ACM patients with

unknown pathogenic gene mutations (9). In addition, this predictive model only incorporated parameters related to the abnormal electroactivity of cardiomyocytes and cardiac fibrosis. We hypothesized that incorporation parameters associated with the overactivity of the cardiac sympathetic system might enhance the value of the prediction model.

SDNN, a parameter of heart rate variability, is regarded as an indicator of cardiac sympathetic system activity. Our previous study revealed that ACM patients exhibited decreased SDNN levels, and SDNN level emerged as an independent predictor of sVT in ACM patients (15). The present study also demonstrated that the SDNN levels in the sVT group were significantly lower than those in the control group. Further analysis using LASSO regression and logistic regression indicated that SDNN levels were independently associated with the occurrence of sVT in patients with ACM. The prediction model incorporating SDNN and these canonical parameters effectively identifies high-risk patients for sVT in individuals with ACM. Furthermore, calibration plots exhibited good agreement between predicted and observed sVT in ACM patients. Therefore, the predictive model in this study may more effectively identify high-risk patients for sVT among individuals with ACM.

The significant clinical benefit of our prediction model lies in its precise quantification of arrhythmic risk tailored to individual ACM patients, offering prognostic insights crucial for guiding clinical decisions regarding prophylactic ICD implantation. Additionally, all parameters included in the prediction model are readily accessible clinical variables, rendering the model user-friendly. It is important to acknowledge that all these parameters in the prediction model might change as ACM progresses. Thus, ACM patients should undergo periodic re-stratification as recommended in a recent expert consensus document.

Limitations

One potential limitation is that all patients in this study were from tertiary medical centers in China and had relatively high incidences of sVT within this cohort (60.5%). The bias in patient selection could potentially impact the overall accuracy of this prediction model when applied to other ACM patients. Secondly, the relatively small sample size of patients used to develop a new prediction model might be associated with selection bias, reduce the statistical power, increase the risk of type II errors, and affect the reliability of the conclusions drawn. Although we have tried to minimize the impact of a small sample size by adding a validation cohort, caution is still required when interpreting and applying the results of our study. Additionally, due to its retrospective design, our study possesses certain drawbacks compared to a prospective study. Moreover, genetics play important roles in the diagnosis of ACM and sVT risk evaluation in ACM (41, 42). However, only a small portion of patients in this study underwent genetic testing, so genetics were not included in the construction of the prediction model. Finally, some *p*-values in this study were close to the significance threshold, raising the potential for type I error. Therefore, further large prospective cohort studies are needed to assess the accuracy and clinical utility of the prediction model in risk stratification for sVT in ACM patients.

In summary, we developed and validated a new prediction model to generate individualized estimates of the risk of incident sVT in ACM patients. Utilizing easily accessible clinical parameters, this model has the potential to accurately identify high-risk patients with sVT, thereby offering significant value for guiding prophylactic ICD implantation in ACM patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the institutional Ethics Committee board of the First Affiliated Hospital of Nanjing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because Given the retrospective nature of the study, individual consent was waived.

Author contributions

BZ: Writing – original draft, Writing – review & editing. XX: Writing – original draft. JY: Methodology, Supervision, Writing – review & editing. YW: Data curation, Formal Analysis, Project administration, Validation, Writing – original draft. JZ: Conceptualization, Formal Analysis, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. XL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Resources, Software, Visualization, Writing – review & editing. BY: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the Key Specialty Construction Project of Shanghai Pudong New Area Health Commission (PWZzk2022-03), the Top-level Clinical Discipline Project of Shanghai Pudong District Grant/Award (PWYgf2021-01), the Health Commission of Shanghai Pudong District (PW2023A-17), and the Science and Technology Plan Project of Health Commission of Jiangxi Province (202310098).

Acknowledgments

We thank all participants in this study and all work staff in the department of cardiology, the First Affiliated Hospital of Nanjing Medical University and Shanghai East Hospital.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2024.1477931/full#supplementary-material>

References

- Corrado D, Link MS, Calkins H. Arrhythmogenic right ventricular cardiomyopathy. *N Engl J Med*. (2017) 376:61–72. doi: 10.1056/NEJMr1509267
- Finocchiaro G, Radaelli D, D'Errico S, Papadakis M, Behr ER, Sharma S, et al. Sudden cardiac death among adolescents in the United Kingdom. *J Am Coll Cardiol*. (2023) 81:1007–17. doi: 10.1016/j.jacc.2023.01.041
- Zhang B, Wu Y, Zhou C, Xie J, Zhang Y, Yang X, et al. Hyperactivation of ATF4/TGF- β 1 signaling contributes to the progressive cardiac fibrosis in arrhythmogenic cardiomyopathy caused by DSG2 variant. *BMC Med*. (2024) 22:361. doi: 10.1186/s12916-024-03593-8
- Zeppenfeld K, Tfelt-Hansen J, de Riva M, Winkel BG, Behr ER, Blom NA, et al. 2022 ESC guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. *Eur Heart J*. (2022) 43:3997–4126. doi: 10.1093/eurheartj/ehac262
- Mariani MV, Pierucci N, Fanisio F, Laviola D, Silvetti G, Piro A, et al. Inherited arrhythmias in the pediatric population: an updated overview. *Medicina (Kaunas)*. (2024) 60:94. doi: 10.3390/medicina60010094
- Towbin JA, McKenna WJ, Abrams DJ, Ackerman MJ, Calkins H, Darrieux FCC, et al. 2019 HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy. *Heart Rhythm*. (2019) 16:e301–72. doi: 10.1016/j.hrthm.2019.05.007
- Wallace R, Calkins H. Risk stratification in arrhythmogenic right ventricular cardiomyopathy. *Arrhythm Electrophysiol Rev*. (2021) 10:26–32. doi: 10.15420/aer.2020.39
- Bosman LP, Te Riele A. Arrhythmogenic right ventricular cardiomyopathy: a focused update on diagnosis and risk stratification. *Heart*. (2022) 108:90–7. doi: 10.1136/heartjnl-2021-319113
- Cadrin-Tourigny J, Bosman LP, Nozza A, Wang W, Tadros R, Bhonsale A, et al. A new prediction model for ventricular arrhythmias in arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J*. (2019) 40:1850–8. doi: 10.1093/eurheartj/ehz103
- Cadrin-Tourigny J, Bosman LP, Nozza A, Wang W, Tadros R, Bhonsale A, et al. A new prediction model for ventricular arrhythmias in arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J*. (2022) 43:e1–9. doi: 10.1093/eurheartj/ehac180
- Protonotarios A, Bariani R, Cappelletto C, Pavlou M, Garcia-Garcia A, Cipriani A, et al. Importance of genotype for risk stratification in arrhythmogenic right ventricular cardiomyopathy using the 2019 ARVC risk calculator. *Eur Heart J*. (2022) 43:3053–67. doi: 10.1093/eurheartj/ehac235
- Aquaro GD, De Luca A, Cappelletto C, Raimondi F, Bianco F, Botto N, et al. Prognostic value of magnetic resonance phenotype in patients with arrhythmogenic right ventricular cardiomyopathy. *J Am Coll Cardiol*. (2020) 75:2753–65. doi: 10.1016/j.jacc.2020.04.023
- Casella M, Gasperetti A, Gaetano F, Busana M, Sommariva E, Catto V, et al. Long-term follow-up analysis of a highly characterized arrhythmogenic cardiomyopathy cohort with classical and non-classical phenotypes—a real-world assessment of a novel prediction model: does the subtype really matter. *Europace*. (2020) 22:797–805. doi: 10.1093/europace/euz352
- Assis FR, Krishnan A, Zhou X, James CA, Murray B, Tichnell C, et al. Cardiac sympathectomy for refractory ventricular tachycardia in arrhythmogenic right ventricular cardiomyopathy. *Heart Rhythm*. (2019) 16:1003–10. doi: 10.1016/j.hrthm.2019.01.019
- Zhang B, Zhou C, Liu J, Zhang J, Jiang C, Tang M, et al. Impaired heart rate variability in patients with arrhythmogenic cardiomyopathy: a multicenter retrospective study in China. *Front Cardiovasc Med*. (2022) 9:1044797. doi: 10.3389/fcvm.2022.1044797
- Marcus FI, McKenna WJ, Sherrill D, Basso C, Baucé B, Bluemke DA, et al. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Eur Heart J*. (2010) 31:1485–96. doi: 10.1093/eurheartj/ehq025
- Zhang B, Yu J, Wu Y, Li X, Xie X, Tao A, et al. The significance of heart rate variability in patients with frequent premature ventricular complex originating from the ventricular outflow tract. *Clin Cardiol*. (2024) 47:e24174. doi: 10.1002/clc.24174
- Bhonsale A, James CA, Tichnell C, Murray B, Gagarin D, Philips B, et al. Incidence and predictors of implantable cardioverter-defibrillator therapy in patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy undergoing implantable cardioverter-defibrillator implantation for primary prevention. *J Am Coll Cardiol*. (2011) 58:1485–96. doi: 10.1016/j.jacc.2011.06.043
- Folino AF, Buja G, Baucé B, Thiene G, dalla Volta S, Nava A. Heart rate variability in arrhythmogenic right ventricular cardiomyopathy correlation with clinical and prognostic features. *Pacing Clin Electrophysiol*. (2002) 25:1285–92. doi: 10.1046/j.1460-9592.2002.01285.x
- Calkins H, Corrado D, Marcus F. Risk stratification in arrhythmogenic right ventricular cardiomyopathy. *Circulation*. (2017) 136:2068–82. doi: 10.1161/CIRCULATIONAHA.117.030792
- Austin KM, Trembley MA, Chandler SF, Sanders SP, Saffitz JE, Abrams DJ, et al. Molecular mechanisms of arrhythmogenic cardiomyopathy. *Nat Rev Cardiol*. (2019) 16:519–37. doi: 10.1038/s41569-019-0200-7
- Segura-Rodriguez D, Bermudez-Jimenez FJ, Carriel V, Lopez-Fernandez S, Gonzalez-Molina M, Oyonarte Ramirez JM, et al. Myocardial fibrosis in arrhythmogenic cardiomyopathy: a genotype-phenotype correlation study. *Eur Heart J Cardiovasc Imaging*. (2020) 21:378–86. doi: 10.1093/ehjci/jez277
- van der Voorn SM, Te Riele A, Basso C, Calkins H, Remme CA, van Veen TAB. Arrhythmogenic cardiomyopathy: pathogenesis, pro-arrhythmic remodelling, and novel approaches for risk stratification and therapy. *Cardiovasc Res*. (2020) 116:1571–84. doi: 10.1093/cvr/cvaa084
- Noorman M, Hakim S, Kessler E, Groeneweg JA, Cox MG, Asimaki A, et al. Remodeling of the cardiac sodium channel, connexin43, and plakoglobin at the intercalated disk in patients with arrhythmogenic cardiomyopathy. *Heart Rhythm*. (2013) 10:412–9. doi: 10.1016/j.hrthm.2012.11.018
- Zhang B, Wu Y, Yang X, Xiang Y, Yang B. Molecular insight into arrhythmogenic cardiomyopathy caused by DSG2 mutations. *Biomed Pharmacother*. (2023) 167:115448. doi: 10.1016/j.biopha.2023.115448
- Krahn AD, Wilde AAM, Calkins H, La Gerche A, Cadrin-Tourigny J, Roberts JD, et al. Arrhythmogenic right ventricular Cardiomyopathy. *JACC Clin Electrophysiol*. (2022) 8:533–53. doi: 10.1016/j.jacep.2021.12.002
- Cerrone M, Montnach J, Lin X, Zhao YT, Zhang M, Agullo-Pascual E, et al. Plakophilin-2 is required for transcription of genes that control calcium cycling and cardiac rhythm. *Nat Commun*. (2017) 8:106. doi: 10.1038/s41467-017-00127-0
- Kim JC, Perez-Hernandez M, Alvarado FJ, Maurya SR, Montnach J, Yin Y, et al. Disruption of $ca(2+)_i$ homeostasis and connexin 43 hemichannel function in the right ventricle precedes overt arrhythmogenic cardiomyopathy in plakophilin-2-deficient mice. *Circulation*. (2019) 140:1015–30. doi: 10.1161/CIRCULATIONAHA.119.039710
- Dusi V, De Ferrari GM. The sympathetic nervous system and arrhythmogenic right ventricular cardiomyopathy: further evidence of a strong tie. *Heart Rhythm*. (2019) 16:1011–2. doi: 10.1016/j.hrthm.2019.02.004
- Wichter T, Hindricks G, Lerch H, Bartenstein P, Borggrefe M, Schober O, et al. Regional myocardial sympathetic dysinnervation in arrhythmogenic right ventricular cardiomyopathy. An analysis using 123I-meta-iodobenzylguanidine scintigraphy. *Circulation*. (1994) 89:667–83. doi: 10.1161/01.CIR.89.2.667
- Maack C, Eschenhagen T, Hamdani N, Heinzel FR, Lyon AR, Manstein DJ, et al. Treatments targeting inotropy. *Eur Heart J*. (2019) 40:3626–44. doi: 10.1093/eurheartj/ehy600
- Herring N, Kalla M, Paterson DJ. The autonomic nervous system and cardiac arrhythmias: current concepts and emerging therapies. *Nat Rev Cardiol*. (2019) 16:707–26. doi: 10.1038/s41569-019-0221-2
- Link MS, Laidlaw D, Polonsky B, Zareba W, McNitt S, Gear K, et al. Ventricular arrhythmias in the north American multidisciplinary study of ARVC: predictors, characteristics, and treatment. *J Am Coll Cardiol*. (2014) 64:119–25. doi: 10.1016/j.jacc.2014.04.035
- Mazzanti A, Ng K, Faragli A, Maragna R, Chiodaroli E, Orphanou N, et al. Arrhythmogenic right ventricular cardiomyopathy: clinical course and predictors of arrhythmic risk. *J Am Coll Cardiol*. (2016) 68:2540–50. doi: 10.1016/j.jacc.2016.09.951
- Corrado D, Calkins H, Link MS, Leoni L, Favale S, Bevilacqua M, et al. Prophylactic implantable defibrillator in patients with arrhythmogenic right ventricular cardiomyopathy/dysplasia and no prior ventricular fibrillation or sustained ventricular tachycardia. *Circulation*. (2010) 122:1144–52. doi: 10.1161/CIRCULATIONAHA.109.913871
- Corrado D, Leoni L, Link MS, Della Bella P, Gaita F, Curnis A, et al. Implantable cardioverter-defibrillator therapy for prevention of sudden death in patients with arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Circulation*. (2003) 108:3084–91. doi: 10.1161/01.CIR.0000103130.33451.D2
- Castillo E. Recurrent syncope in a patient with arrhythmogenic right ventricular cardiomyopathy. *Cureus*. (2023) 15:e45850. doi: 10.7759/cureus.45850
- Corrado D, Anastakis A, Basso C, Baucé B, Blomstrom-Lundqvist C, Bucciarelli-Ducci C, et al. Proposed diagnostic criteria for arrhythmogenic cardiomyopathy: european task force consensus report. *Int J Cardiol*. (2024) 395:131447. doi: 10.1016/j.ijcard.2023.131447
- Efremidis M, Vlachos K, Kyriakopoulou M, Mililis P, Martin CA, Bazoukis G, et al. The RV(1)-V(3) transition ratio: a novel electrocardiographic criterion for the differentiation of right versus left outflow tract premature ventricular complexes. *Heart Rhythm O2*. (2021) 2:521–8. doi: 10.1016/j.hroo.2021.07.009
- Laredo M, Tovia-Brodie O, Milman A, Michowitz Y, Roudijk RW, Peretto G, et al. Electrocardiographic findings in patients with arrhythmogenic cardiomyopathy and right bundle branch block ventricular tachycardia. *Europace*. (2023) 25:1025–34. doi: 10.1093/europace/evac267
- Casian M, Papadakis M, Jurcut R. Arrhythmogenic right ventricular cardiomyopathies (ARVC): diagnostic challenges from imaging to genetics. *Kardiol Pol*. (2024) 82:1059–70. doi: 10.33963/v.phj.102391
- Hoorntje ET, Burns C, Marsili L, Corden B, Parikh VN, Te Meerman GJ, et al. Variant location is a novel risk factor for individuals with arrhythmogenic cardiomyopathy due to a desmoplakin (DSP) truncating variant. *Circ Genom Precis Med*. (2023) 16:e003672. doi: 10.1161/CIRCGEN.121.003672



OPEN ACCESS

EDITED BY

Evangelos Oikonomou,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Pantaleimon Pantelidis,
National and Kapodistrian University of Athens,
Greece
Mingfeng Jiang,
Zhejiang Sci-Tech University, China

*CORRESPONDENCE

Qiang Wu
✉ wuqiang@gz5055.com

RECEIVED 06 August 2024

ACCEPTED 12 December 2024

PUBLISHED 06 January 2025

CITATION

He Y, Zhou Y, Qian Y, Liu J, Zhang J, Liu D and
Wu Q (2025) Cardioattentionnet: advancing
ECG beat characterization with a high-
accuracy and portable deep learning model.
Front. Cardiovasc. Med. 11:1473482.
doi: 10.3389/fcvm.2024.1473482

COPYRIGHT

© 2025 He, Zhou, Qian, Liu, Zhang, Liu and
Wu. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Cardioattentionnet: advancing ECG beat characterization with a high-accuracy and portable deep learning model

Youfu He^{1,2,3}, Yu Zhou^{2,3}, Yu Qian⁴, Jingjie Liu⁵, Jinyan Zhang⁶,
Debin Liu⁷ and Qiang Wu^{2,3*}

¹Medical College, Guizhou University, Guiyang, Guizhou, China, ²Department of Cardiology, Guizhou Provincial People's Hospital, Guiyang, Guizhou, China, ³Department of Cardiology, Guizhou Provincial Cardiovascular Disease Clinical Medicine Research Center, Guiyang, Guizhou, China, ⁴Department of Cardiology, The Second Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou, China, ⁵Department of Cardiology, First Affiliated Hospital of Dalian Medical University, Liaoyang, Liaoning, China, ⁶School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, ⁷Department of Cardiology, The Second People's Hospital of Shantou, Shantou, Guangdong, China

Introduction: The risk of mortality associated with cardiac arrhythmias is considerable, and their diagnosis presents significant challenges, often resulting in misdiagnosis. This situation highlights the necessity for an automated, efficient, and real-time detection method aimed at enhancing diagnostic accuracy and improving patient outcomes.

Methods: The present study is centered on the development of a portable deep learning model for the detection of arrhythmias via electrocardiogram (ECG) signals, referred to as CardioAttentionNet (CANet). CANet integrates Bi-directional Long Short-Term Memory (BiLSTM) networks, Multi-head Attention mechanisms, and Depthwise Separable Convolution, thereby facilitating its application in portable devices for early diagnosis. The architecture of CANet allows for effective processing of extended ECG patterns and detailed feature extraction without a substantial increase in model size.

Results: Empirical results indicate that CANet outperformed traditional models in terms of predictive performance and stability, as confirmed by comprehensive cross-validation. The model demonstrated exceptional capabilities in detecting cardiac arrhythmias, surpassing existing models in both cross-validation and external testing scenarios. Specifically, CANet achieved high accuracy in classifying various arrhythmic events, with the following accuracies reported for different categories: Normal ($97.37 \pm 0.30\%$), Supraventricular ($98.09 \pm 0.25\%$), Ventricular ($92.92 \pm 0.09\%$), Atrial Fibrillation ($99.07 \pm 0.13\%$), and Unclassified arrhythmias ($99.68 \pm 0.06\%$). In external evaluations, CANet attained an average accuracy of 94.41%, with the area under the curve (AUC) for each category exceeding 99%, thereby demonstrating its substantial clinical applicability and significant advancements over traditional models.

Discussion: The deep learning model proposed in this study has the potential to enhance the accuracy of early diagnosis for various types of arrhythmias. Looking ahead, this technology is anticipated to provide improved medical services for patients with heart disease through continuous, non-invasive monitoring and timely intervention.

KEYWORDS

cardiac arrhythmias, electrocardiogram, portable deep learning model, transformer model, Long Short-Term Memory, MobileNet

1 Introduction

Cardiac arrhythmias, resulting from disruptions in the heart's electrical signals, frequently occur in clinical settings. Cardiac arrhythmias, which manifest in multiple forms with varying mechanisms and severities, are a major cause of morbidity and mortality (1). The China Hypertension Survey conducted from 2012 to 2015 reported a prevalence rate of 0.7% for atrial fibrillation in individuals aged 35 years and older, with 34.0% of these cases being undetected by the patients themselves (2). In the United States, most sudden cardiac deaths (SCDs) were due to ventricular tachyarrhythmias, making up 80% of these incidents (3). Bradyarrhythmias and conduction disorders can precipitate syncope, sudden cardiac death, and symptoms including fatigue and exercise intolerance due to intermittent heart rate inadequacy. However, these conditions can be difficult to identify (4).

Clinical screening for arrhythmia predominantly relies on patients presenting to healthcare facilities after the onset of symptoms. Under the supervision of specialized electrocardiographers or cardiologists, physicians are tasked with diagnosing arrhythmias. Furthermore, some patients may be identified as having various forms of arrhythmia during routine physical examinations (5). However, both detection methods necessitate a high degree of professionalism from healthcare providers. A cross-sectional study indicated that primary care physicians exhibited a misdiagnosis rate of 23% for abnormal electrocardiograms (6). It is noteworthy that, despite their expertise in cardiovascular medicine, the average diagnostic accuracy for ventricular tachycardia is only 78.4% (7). The large-scale training of specialized cardiologists is particularly impractical in rural regions worldwide, especially in under-developed countries. The advancement of artificial intelligence technology has facilitated the increased application of deep learning techniques for arrhythmia detection. Nonetheless, the integration of AI and deep learning into electrocardiography (ECG) detection poses distinct challenges due to the specialized nature of ECG interpretation.

Frameworks for interpreting electrocardiograms (ECGs), including parameters such as heart rate, rhythm, cardiac axis, intervals, and ventricular activity, have been developed to classify and identify various cardiac disorders in contemporary research. For instance, Attia et al. introduced a novel algorithm utilizing convolutional neural networks to predict paroxysmal atrial fibrillation (AF) in patients, achieving an overall accuracy of 83.3% based on benign, normal sinus rhythm ECGs (8). Similarly, Khurshid et al. (9) demonstrated that ECG *P*-waves significantly impacted the predictions made by artificial intelligence (AI) models during the training of convolutional neural networks. Their study employed a combined analysis of AI and clinical risk factor models for AF, which enhanced the functionality and accuracy of ECG-AI systems. Despite the relative robustness of these systems, the complex nature of the cardiac conduction system results in numerous ECG graphs exhibiting morphological characteristics that are not readily

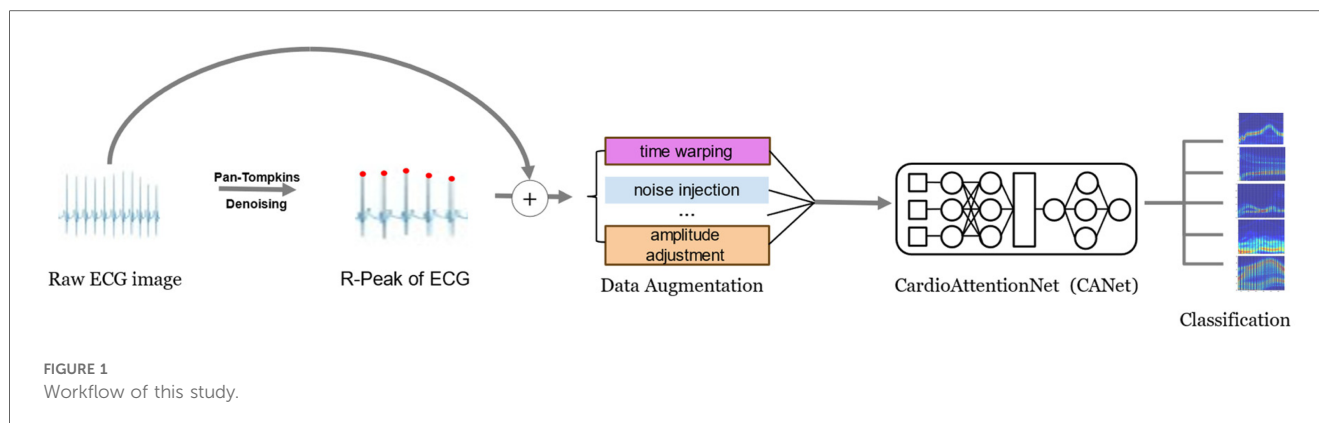
discernible to the human eye, complicating the identification of local morphology and inter-beat relationships. Furthermore, there remains a lack of comprehensive evaluation regarding the efficacy of end-to-end deep learning methods in classifying a diverse array of diagnoses from raw ECG data (7). Previous research has focused on specific components of the ECG processing workflow, such as localized noise reduction (10) and feature extraction (9), or has been limited to diagnosing specific rhythms, primarily AF and myocardial infarction (MI) (9, 11). This narrow focus contributes to suboptimal accuracy rates in ECG-AI applications and restricts the advancement of AI and deep learning methodologies in this domain (12).

In this study, we address the evolving trends in deep learning and the challenges associated with cardiac arrhythmia detection by introducing the innovative CardioAttentionNet (CANet) model. This model is specifically designed to facilitate portable and accurate arrhythmia detection, thereby enabling early discovery and diagnosis, which can significantly reduce potential harm to patients. CANet is constructed upon the Long Short-Term Memory (LSTM) network and incorporates the Multi-Head Attention mechanism derived from the Transformer model, as well as Depthwise Separable Convolution from MobileNet. A comparative analysis with traditional models indicates that the Bi-directional LSTM (BiLSTM) performs effectively in arrhythmia detection due to its capability to process and retain long-term sequential information, which is essential for analyzing extended rhythm patterns and cyclic variations in electrocardiogram (ECG) signals. However, BiLSTM is limited in its ability to simultaneously focus on multiple features and capture finer details. To overcome this limitation, we have integrated Multi-Head Attention, which allows the model to concentrate on various aspects of ECG signals, in conjunction with Depthwise Separable Convolution for enhanced feature extraction. When compared to baseline models, CANet exhibits superior accuracy and predictive performance, effectively identifying a range of cardiac arrhythmias. The workflow of this study is depicted in Figure 1.

2 Methods

2.1 Data collection

In this study, we utilized the widely recognized MIT-BIH Arrhythmia Database (13, 14) for our experimental analyses. This database was particularly well-suited for training deep neural networks due to its extensive collection of samples. It consisted of 109,446 electrocardiogram (ECG) recordings, each sampled at a frequency of 360 Hz, which includes both normal heart rhythms and various arrhythmic conditions. Notably, the database segmented these ECG signals into individual beats, each representing a distinct cardiac cycle. Following the standard clinical classification, the heartbeats were categorized into five types: Normal beats (N), Fusion beats (F), Supraventricular ectopic beats (S), Ventricular ectopic beats (V), and Unknown



beats (Q). In our research, we adhered to this classification scheme for the identification and analysis of arrhythmias within the ECG signals.

2.2 Data preprocessing

Despite the pre-segmentation of electrocardiogram (ECG) signals in the MIT-BIH Arrhythmia Database, these signals continued to exhibit a considerable amount of noise, including electromyographic interference, motion artifacts, and baseline drift, among other factors. Such noise could negatively impact the quality of heart rate signals, thereby hindering the precise extraction of features. Consequently, the preprocessing of these signals was imperative. In the domain of ECG signal preprocessing, wavelet transform was a widely utilized method (15). This technique facilitated the transformation of long-duration time-domain signals into time-frequency representations, thereby elucidating the frequency domain components at each time point and capturing more localized features. During the wavelet transform process, we primarily employed the Mallat algorithm (16). This method systematically decomposed the signal into “low-frequency approximations” and “high-frequency details,” which allowed for the effective removal of frequencies associated with electromyographic noise (20–5,000 Hz) and baseline drift (1 Hz). Following the noise reduction and signal restoration, we achieved ECG signals with improved accuracy and clarity, thereby establishing a robust foundation for subsequent data augmentation processes (17).

Recognizing the disparity in the MIT-BIH Arrhythmia Database, with a significant predominance of normal heart rate samples over those exhibiting arrhythmias, we employed data augmentation methods in the preprocessing stage. This approach ensured a balanced representation of all classes in the dataset and bolstered the model’s ability to detect abnormal patterns, thus enhancing its overall generalizability. Specifically, we employed data augmentation methods such as time warping and noise injection to create a balanced dataset. These techniques markedly improved the performance of our model, effectively addressing the challenges associated with data imbalance. The

adjusted proportions of the different arrhythmia categories in the dataset, both prior to and following data augmentation, were depicted in Figure 2. Experimental results demonstrated that these data augmentation strategies significantly bolstered the model’s performance and successfully mitigated the issues related to data imbalance.

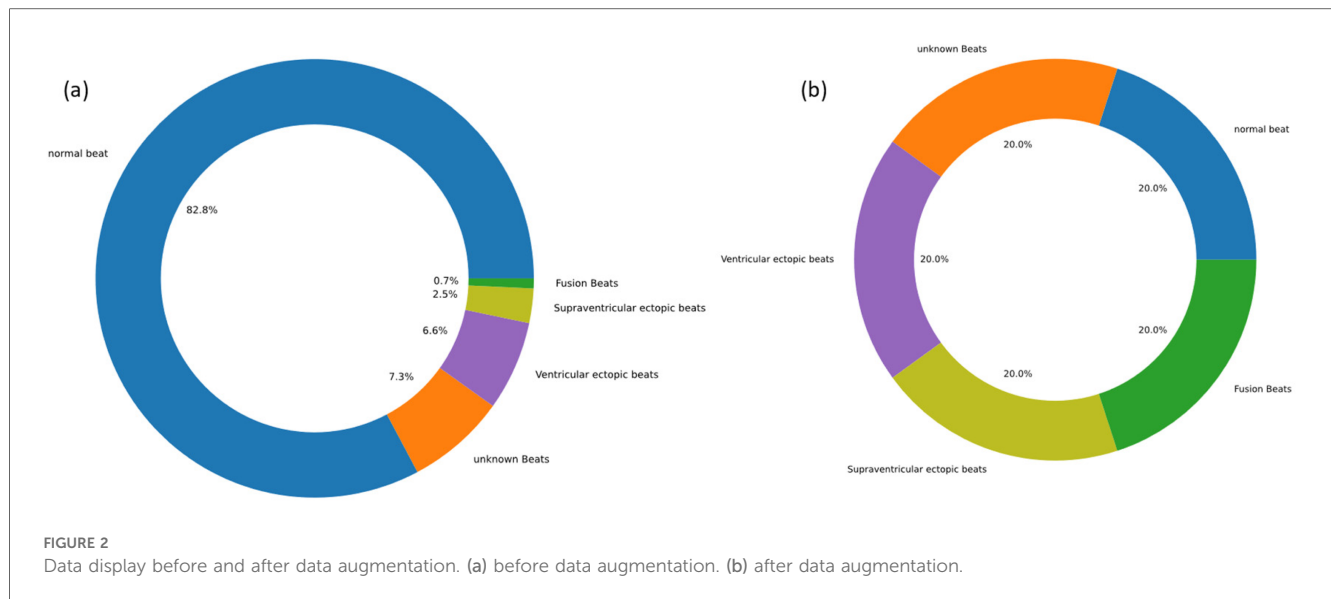
2.3 Feature extraction

In this study, we selected segmented electrocardiogram (ECG) signals as the primary input, supplemented by each heartbeat interval (R-R interval, RRI) as a parallel input, to establish comprehensive indicators that reflect the key characteristics of cardiac arrhythmias. To enhance accuracy in subsequent deep learning processes, we employed Gaussian filtering to improve data contrast. Although Gaussian filters are predominantly utilized in digital image processing, they are equally effective in enhancing the signal-to-noise ratio, filtering out noise, and ensuring smoother ECG signals (18). Given the waveform deformations frequently observed in ECGs during arrhythmias, we implemented R-wave detection to delineate each individual heartbeat. In our methodology, the Pan-Tompkins algorithm (19, 20) was initially employed to detect R-waves in the denoised ECG signals on a per-minute basis. Subsequently, we calculated the intervals between adjacent R-waves, yielding RRI in milliseconds as a parallel input to the ECG signals, thereby providing a more intuitive representation of the characteristics of cardiac arrhythmias.

2.4 Model construction

2.4.1 Cardioattentionnet (CANet)

The advancements in deep learning have significantly enhanced its application in the detection of cardiac arrhythmias, particularly through the utilization of models such as Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNN). These models, especially LSTM, have demonstrated superior recognition accuracy compared to



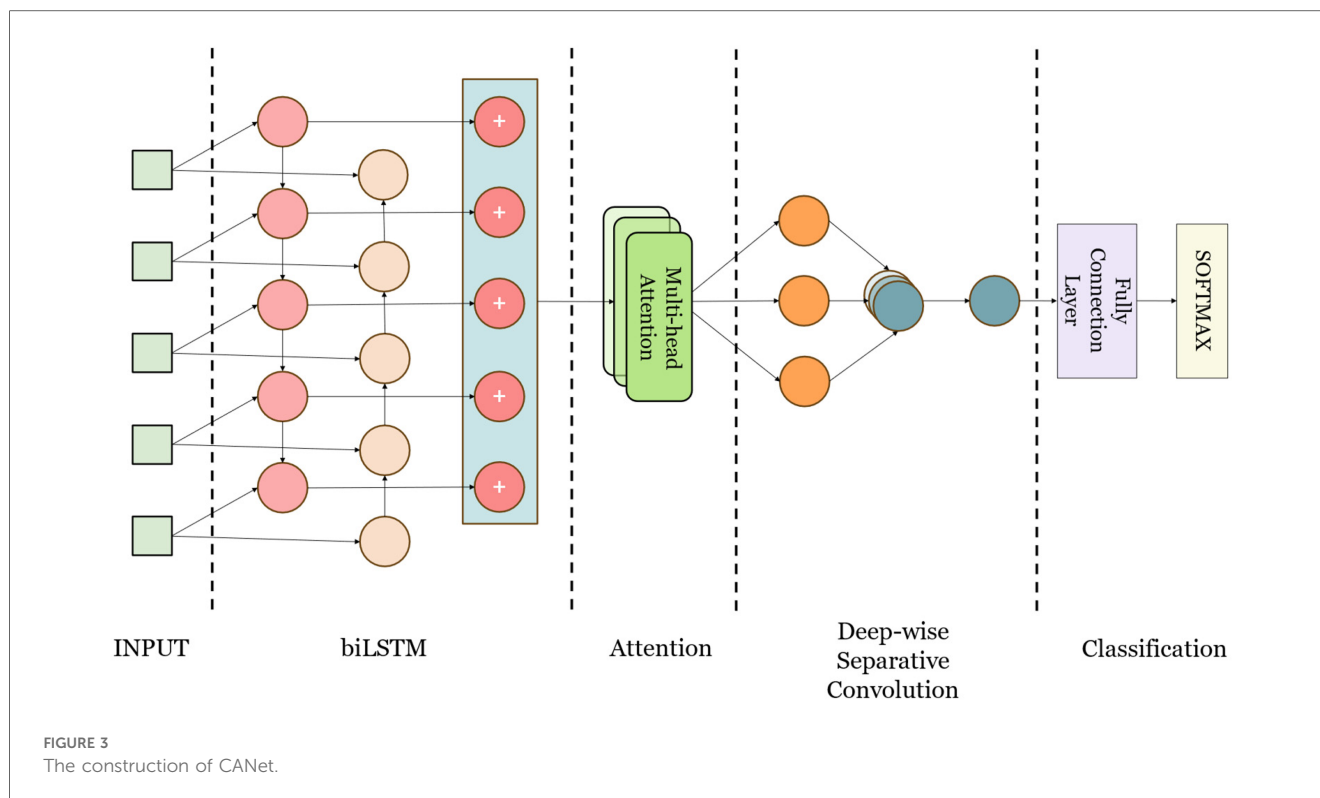
traditional methods (21). LSTM, which represented an improvement over RNN, effectively addresses challenges associated with gradient vanishing and long-term dependencies, thereby enhancing the accuracy of deep learning applications. The efficacy of RNN and LSTM in analyzing time-dependent data was particularly noteworthy, as evidenced by the work of Salloum, Ronald, and Kuo, C.-C. Jay, who reported elevated arrhythmia detection rates using RNN without the necessity for prior feature extraction (22). Furthermore, Hannun et al. (7) introduced a Deep Neural Network (DNN) for classification purposes, underscoring the high sensitivity and predictive capabilities of DNNs. Nonetheless, concerns regarding the accuracy of outputs generated without preprocessing input data persist, highlighting the necessity for a careful balance between accuracy and computational efficiency (21). RNN models retain previous ECG information, influencing current inputs. LSTM units loop information across time steps, creating internal feedback that helps the network understand time and learn temporal dynamics in the data. With these properties, Currently, many models have been using RNN and LSTM models to achieve demonstrated accuracy in arrhythmia detection, including Sumanta et al.'s ELM-RNN model (ECG signal classification and arrhythmia detection using ELM-RNN) and Shu Liu Oh's CNN-LSTM model (Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats).

This study presents CANet, a hybrid deep learning model developed for the detection of cardiac arrhythmias. CANet integrates Bidirectional Long Short-Term Memory (BiLSTM), Multihead Attention, and Depthwise Separable Convolution to enhance generalization, predictive performance, and structural robustness, while simultaneously minimizing bias and variance. The design of the model leverages the strengths of various machine learning techniques, amalgamating them into a cohesive framework, as illustrated in Figure 3.

CANet architecture initiates with Bidirectional Long Short-Term Memory (BiLSTM) processing, which enhances temporal dependencies by integrating both past and future data, thereby rendering it appropriate for electrocardiogram (ECG) signal analysis. This configuration facilitates the retention of long-term dependencies, consequently improving classification accuracy. Subsequently, the Multi-Head Attention layer, derived from Transformer models and adapted for image and time series applications, allocates attention to various features concurrently, thereby mitigating the issue of feature neglect (23). This aspect is particularly significant in ECG signal analysis, as it addresses the challenges associated with feature focus in BiLSTM.

The model utilizes Depthwise Separable Convolution, which demonstrates greater efficiency compared to standard convolutions in terms of both parameter count and computational workload, thereby improving training speed (24). This architecture is particularly effective in extracting local time-domain features from electrocardiogram (ECG) signals, which is essential for the identification of abnormal ECG characteristics.

Following comprehensive testing and validation, CANet demonstrates considerable advantages in the detection of cardiac arrhythmias, outperforming conventional models. The architecture incorporates Bidirectional Long Short-Term Memory (BiLSTM) networks, attention mechanisms, and convolutional networks, thereby effectively capturing anomalies in electrocardiogram (ECG) signals from various perspectives. Cardio-AttentionNet is specifically designed to perform efficiently in data-scarce environments, where smaller datasets are often a reality. Through its lightweight architecture and attention-based feature selection, the model is capable of leveraging limited data effectively, as demonstrated by its strong performance on the MIT-BIH Arrhythmia Database. External evaluations and five-fold cross-validation further substantiate CANet's exceptional performance in arrhythmia detection.



2.4.2 Baseline models

To conduct an objective evaluation of the proposed CANet model, this study established several baseline models for comparative analysis to determine the efficacy and advantages of CANet. The selected baseline models for comparison included Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Units (GRU). These models were chosen due to their prevalent application in deep learning and their significant contributions to research in heart rate detection.

Convolutional Neural Networks (CNNs), a prevalent architecture in deep learning, were recognized for their efficiency in extracting signal features via convolutional operations. This model has gained prominence in the domains of signal classification and segmentation, particularly in the detection of electrocardiogram (ECG) signals, where it has demonstrated a significant impact (25, 26).

Recurrent Neural Networks (RNNs) were acknowledged for their strong model fitting and predictive capabilities when dealing with sequential data, and they have demonstrated significant effectiveness in the detection of arrhythmias. In contrast to conventional deep learning architectures, the hidden layers of RNNs incorporated memory functions that facilitated the retention and utilization of previous information. This characteristic was particularly vital in the analysis of electrocardiogram (ECG) signals (27, 28).

Long Short-Term Memory (LSTM), an extension of Recurrent Neural Networks (RNN), integrates “gates” at each unit (29), thereby enhancing the model’s memory capabilities and

providing significant advantages in the processing of longer sequences. LSTM has become a widely utilized model in the detection of arrhythmias (30).

The Gated Recurrent Unit (GRU), a variant of the Recurrent Neural Network (RNN) developed to mitigate challenges associated with long-term memory and gradient descent, features a reduced number of “gates” in comparison to the Long Short-Term Memory (LSTM) architecture, thereby improving computational efficiency (31). Furthermore, the GRU has demonstrated significant efficacy in the analysis of electrocardiogram (ECG) signals and related applications.

Bidirectional Long Short-Term Memory (BiLSTM), an enhancement of Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, incorporates a mechanism for the propagation of data in both forward and backward directions. This capability allows BiLSTM to integrate information from both temporal contexts, thereby improving its predictive performance compared to traditional RNN and LSTM architectures (32). Consequently, BiLSTM has attracted significant research interest within the domain of electrocardiogram (ECG) analysis (33–35).

2.5 Model evaluation

In order to conduct a thorough and objective assessment of the performance of the proposed CANet model, this study utilized standard evaluation metrics, including Accuracy (ACC), Precision (PRE), Recall (REC), F1-Score (F1), and the Receiver Operating Characteristic (ROC) curve. These metrics were derived from the values of True Positives (TP), True Negatives (TN), False

Positives (FP), and False Negatives (FN). The specific formulas for these metrics were outlined as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$PRE = \frac{TP}{TP + FP} \quad (2)$$

$$REC = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times PRE \times REC}{PRE + REC} \quad (4)$$

where:

- ACC represents the ratio of correctly classified samples to the total number of samples, providing a direct measure of model performance.
- PRE reflects the proportion of actual positive samples that are predicted as positive by the model, indicating the precision of the model. High precision is crucial in applications such as healthcare, where it denotes a lower rate of misclassification.
- REC indicates the proportion of actual positive samples correctly identified by the model, representing the model's coverage. A high recall signifies that the model captures most of the real positive cases.
- The F1 score, the harmonic mean of precision and recall, is an integrated metric, especially important when there is an imbalance between positive and negative samples.

Additionally, the ROC curve graphically represents the performance of a binary classifier, showing the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at various threshold settings. Ideally, the best classifier's ROC curve approaches the upper left corner of the graph. The Area Under the Curve (AUC) measures the model's overall ability to distinguish between positive and negative samples. An AUC value close to 1 indicates superior model performance.

3 Results

3.1 Experiment setup

In this study, comprehensive parameter tuning and optimization were conducted on all developed models to ensure the objectivity and accuracy of the results. To maintain consistency, identical parameter settings were employed throughout the five-fold cross-validation and external testing processes. Iterative experimentation demonstrated that the models converged effectively without overfitting when the number of epochs was set to 30. Specifically, for our CANet model, the learning rate and batch size were fixed at 0.0001, while parameter adjustments were made based on gradient thresholding to enhance model performance.

All experiments were conducted on a platform running Windows 11 Professional operating system, with Python 3.10.9 as the runtime environment. Regarding software libraries, Pytorch 2.0.1 + cu117, Scikit-learn, Sklearn 0.0.post1, scipy 1.10.0, and other mathematical libraries were utilized to support the construction of the model structure and the validation of results. The hardware configuration included an Intel Core i7 10750H processor (base frequency of 2.6 GHz, turbo frequency up to 5 GHz, 6 cores/12 threads) and an NVIDIA GeForce GTX 1080Ti graphics card (8 GB memory capacity, 128-bit memory bus width).

3.2 Result of five-fold cross-validation

To comprehensively and objectively evaluate the performance of the proposed CANet model, this study utilized the five-fold cross-validation method for dataset division and conducted rigorous testing of the model. The mean values of the model test results were used to reduce randomness and improve generalization capability. After 10 training epochs, optimal performance for CANet was observed with a learning rate of 0.0001, which helps to ensure that the model's weights are updated in small steps, preventing large oscillations or divergence in the loss function. To provides a good balance between accurate gradient estimation and computational efficiency, the batch size is 64, without overfitting. The final training iteration graph is depicted in [Figure 4a](#).

During the five-fold cross-validation, CANet showed a performance higher than all other models. In the accuracy scale, this model shown a stable ability of prediction, which was the only one that recognized all five heartbeats with a accuracy higher than 95%. Moreover, it outperformed all baseline models across all evaluation metrics (ACC, PRE, REC and F1), showcasing exceptional predictive performance.

These results highlight CANet's high accuracy in recognizing five different types of heartbeats. The corresponding ROC curves and confusion matrices are shown in [Figures 4b,c](#).

Compared to the five baseline models, CANet showed more favorable results in the cross-validation. The baseline models' ACC, PRE, REC, and F1 scores were lower than CANet's, which is shown in the [Supplementary Table S1](#).

Among them, the GRU model performed the best, which mean ACC reached 97%, but its precision was significantly lower than CANet. The BiLSTM model slightly trailed GRU in classification accuracy. The performance of the remaining models declined further across several metrics. Overall, traditional models fell short in various metrics compared to CANet, which demonstrated a clear advantage in the cross-validation. The ROC curves for the baseline models are presented in [Figures 4c–g](#).

3.3 Result of external validation

To evaluate the performance of the newly developed CANet model, an external test using an unseen sample set was

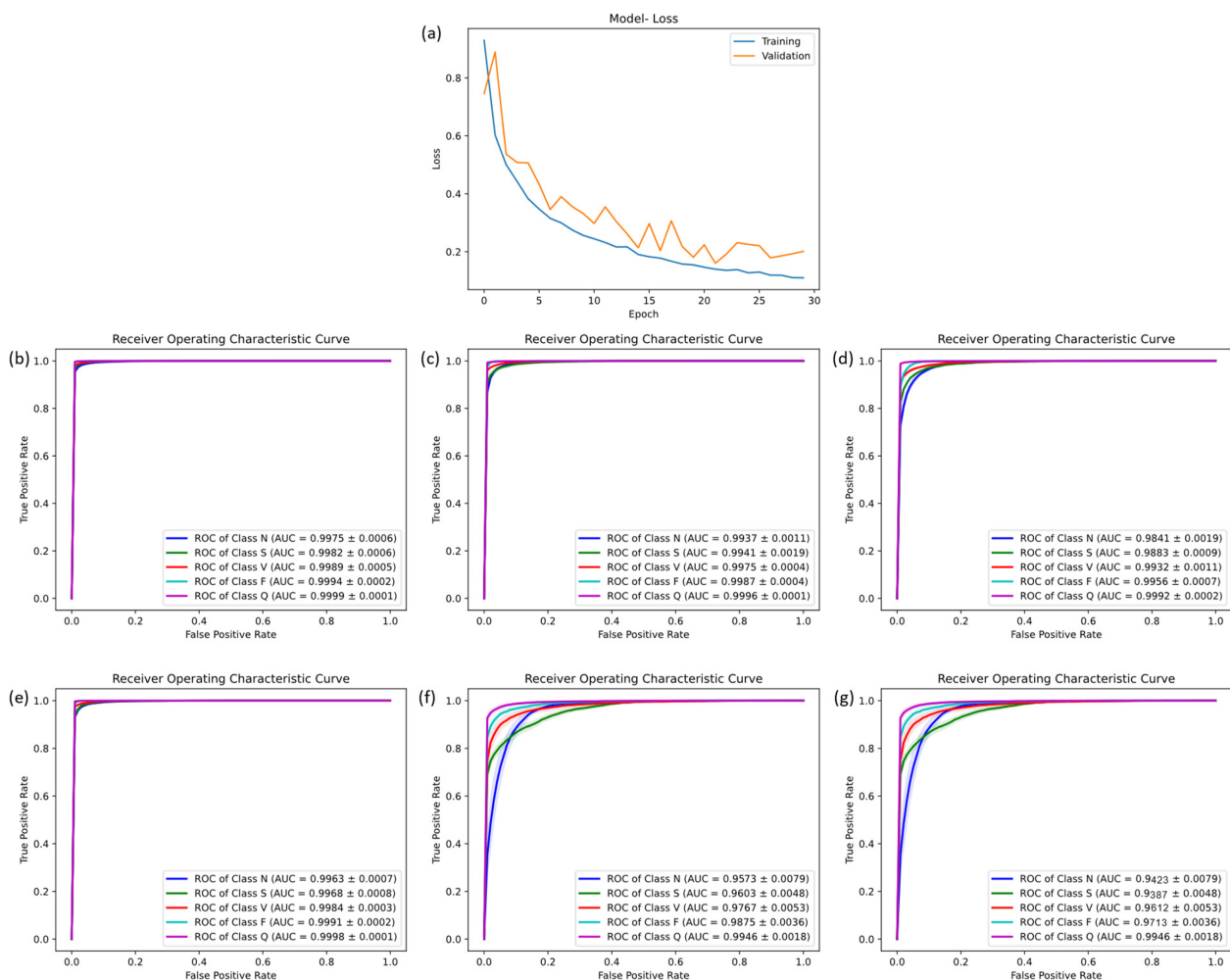


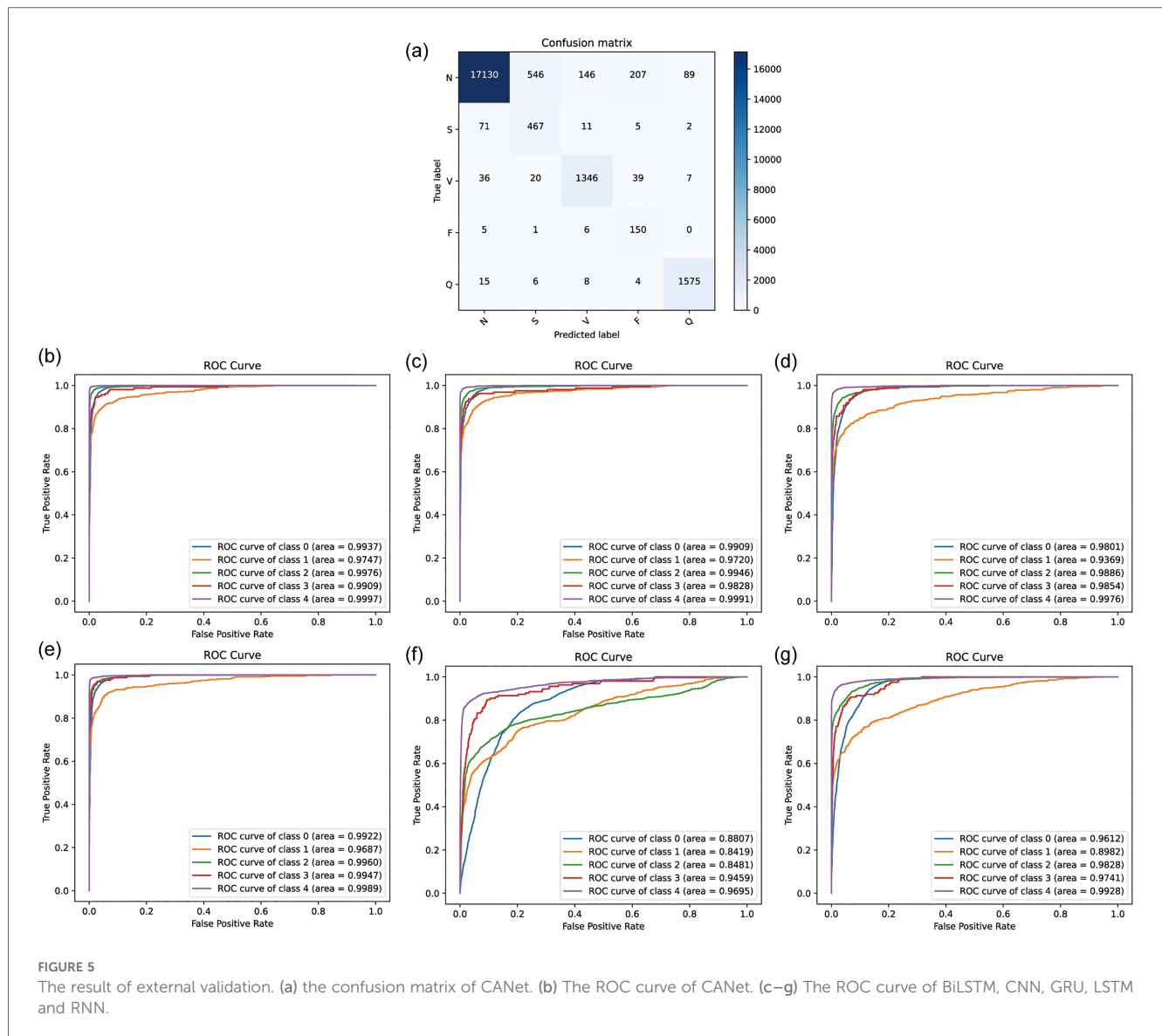
FIGURE 4

The result of five-fold cross-validation. (a) the training flow of CANet. (b) The ROC curve of CANet. (c–g) The ROC curve of BiLSTM, CNN, GRU, LSTM and RNN.

conducted to assess its capability in handling unseen data. The results demonstrate that CANet performed exceptionally well, with average ACC, PRE, REC, and F1 of 94.41%, 92.41%, 92.41%, and 79.02%, respectively. Notably, the model accurately identified [S] and [Q] type, achieving precision rates of 99.26% and 99.22%, respectively. For [N] type, CANet consistently showed reliable performance across all metrics, with ACC, PRE, REC, and F1 at 94.91%, 99.26%, 94.55%, and 96.85%, respectively. Additionally, the model was trained and tested using data prior to data augmentation, yielding average ACC and PRE of 93.27% and 87.23%, respectively. Due to imbalances in the dataset, a significant number of [Q] type predictions were incorrectly classified as [N], markedly reducing precision. These findings underscore the significant improvement in model performance facilitated by data augmentation. Overall, CANet maintained high accuracy and precision across different types of heartbeats without significant performance variation. The corresponding confusion matrix and ROC curve are presented in Figures 5a,b.

In external testing of baseline models, the GRU model exhibited the highest accuracy, with mean ACC, PRE, REC, and F1 of 93.76%, 73.06%, 93.14%, and 79.39%, respectively. Despite its high recognition rate, GRU's performance across all evaluation metrics was inferior to CANet, especially in PRE and F1. The performance of all models in the external test is detailed in [Supplementary Table S2](#).

Other baseline models demonstrated lower precision, accuracy, recall, and F1-scores. Additionally, GRU showed lower precision in recognizing [S] and [V] types, with rates of only 38.38% and 41.86%, respectively. The accuracy of BiLSTM was slightly lower than GRU, at 91.74%, with a precision of only 67.43%. The remaining models (LSTM, CNN, RNN) showed poorer recognition results, with lower accuracy, precision, recall, and F1. The results indicate that baseline models were inferior to CANet in terms of accuracy and adaptability to the dataset, displaying a clear imbalance in recognizing different types of heartbeats. Additionally, the ROC curves of the baseline models are illustrated in Figures 5b–g. Analysis of these curves reveals that



the CANet model surpasses all baseline models in AUC values for each classification. Apart from the [S] category, which has an AUC of 97.47%, all other classifications have AUCs exceeding 99%, demonstrating exceptional clinical predictive value. These results confirm the superior predictive performance of the CANet model. Given its lightweight design, the model is poised for widespread deployment in clinical electrocardiographic devices and portable wearable devices, facilitating precise and real-time detection of cardiac arrhythmias.

We measured the inference time for the model to process a single ECG signal on a standard hardware platform (Intel Core i7 CPU with 16 GB RAM). The average inference time was found to be approximately 56.7 ms, which is well within the typical requirement for real-time detection. This result is consistent with the claim that the model can be used for real-time arrhythmia detection. The model was found to require approximately 30 MB of memory during inference. This low memory footprint demonstrates the portability of

CardioAttentionNet, making it feasible for deployment on resource-constrained devices, such as wearable ECG monitors or mobile health applications.

4 Discussion

In this study, we proposed an innovative fusion model referred to as CANet, which integrates Bidirectional Long Short-Term Memory (BiLSTM), Multihead Attention, and Depthwise Separable Convolution. This model was specifically designed to achieve high accuracy and robustness while maintaining a lightweight architecture. CANet is employed for the classification of electrocardiogram (ECG) signals, with the objective of distinguishing abnormal heartbeats from normal cardiac rhythms, thereby facilitating the detection of cardiac arrhythmias. The results indicated that CANet exhibits remarkable recognition capabilities for five clinically distinct types of heartbeats.

During five-fold cross-validation, CANet achieved high average precision, with 95% confidence intervals for the various heartbeat types as follows: Normal (N) $97.92 \pm 0.33\%$, Supraventricular (S) $98.58 \pm 0.21\%$, Ventricular (V) $99.13 \pm 0.13\%$, Fusion (F) $99.30 \pm 0.17\%$, and QRS (Q) $99.74 \pm 0.5\%$. In external testing, CANet continued to demonstrate excellent performance, attaining average accuracy (ACC), precision (PRE), recall (REC), and F1 scores of 94.91%, 99.26%, 94.55%, and 96.85%, respectively.

Although the CANet model demonstrated superiority across multiple tasks in this study, we also acknowledge that other models retain advantages in specific scenarios. The GRU model, in particular, excels in handling time-series data, especially in situations characterized by strong short-term dependencies. Due to its streamlined architecture and lower computational complexity, the GRU may offer higher efficiency in environments with limited resources or computing power. In contrast, while CANet boasts more robust performance, its complexity may render it less suitable for scenarios sensitive to computational speed or resource demands. Moreover, despite CANet's demonstrated higher accuracy and robustness in our investigated tasks, lightweight models such as the GRU might still be viable choices in tasks or with data features where these characteristics are prominent. These models could have advantages in applications requiring real-time responses or operating under resource constraints. Therefore, the selection of models in practical applications should consider the specific task requirements, computational resources, and data characteristics comprehensively.

The exemplary performance of CANet can be attributed to the synergistic combination of its architectural design and predictive capabilities. The integration of LSTM, Multi-Head Attention, and Depthwise Separative Convolution enables effective processing and memory of long-term sequential data, identification of cyclic patterns in ECG signals, and simultaneous focus on multiple features. Additionally, the temporal nature of cardiac signals offers an ideal setting for the BiLSTM component of the model, yielding more precise and efficient recognition capabilities. Through this optimized design, we have significantly enhanced the model's performance while minimizing its parameters, thus avoiding excessive computational load on the hosting devices. As a result, the model holds potential for future deployment in portable devices for real-time and precise detection and classification of cardiac arrhythmias. These combined elements contribute to the outstanding performance of CANet, showing potential for application in wearable devices and clinical practice for early arrhythmia diagnosis, thereby mitigating its impact on health.

To illustrate the model's feature extraction and effective recognition of ECG signals, a heat activation map is presented as a visualization of the model output in Figure 6. Specifically, for Normal beats (N), the heatmap shows significant activation during the QRS complex, indicating accurate identification of this critical interval with reduced activation during the T wave, consistent with normal electrophysiological characteristics. For Fusion beats (F), the activation pattern is uniformly distributed during the QRS complex and P wave, reflecting variations in cardiac signals in these regions, corresponding to the clinically observed shortened interval between the P wave and QRS complex. For Supraventricular ectopic beats (S), the heatmap exhibits heightened

activity at the onset of the QRS complex, suggesting the model's sensitivity to changes in QRS morphology typical of supraventricular ectopy, aligning with the clinical presentation of abnormal P waves and altered QRS morphology. For Ventricular ectopic beats (V), the clinical presentation of wide and abnormal QRS complexes without preceding P waves is mirrored by intense activity throughout the QRS interval on the heatmap, consistent with the characteristics of ventricular ectopy. For Unknown beats (Q), significant activation is shown on the T wave following the QRS complex, indicating that the model can effectively capture features of abnormal waveforms when dealing with uncertain categories. This visual representation demonstrates CANet's ability to differentiate between various heart rate variabilities, emphasizing the model's interpretability and potential to enhance trust among medical professionals and patients.

The electrocardiogram (ECG) serves as an essential tool in the routine practice of clinical medicine, with over 300 million ECGs conducted worldwide each year (36). Arrhythmias are highly prevalent; however, the challenges associated with monitoring certain malignant arrhythmias contribute to elevated mortality rates. The enhanced accuracy in diagnosing cardiac arrhythmias is primarily attributed to the expertise of clinicians, a situation that is increasingly untenable given the current global shortage of specialized cardiologists. For instance, China has only 4.8 cardiologists per 100,000 individuals (37), while Japan is reported to have approximately 14,000 specialized cardiologists (38), figures that are grossly insufficient relative to their large populations. In response to these challenges, wearable devices incorporating deep learning technologies are being increasingly integrated into daily life (39). Research indicates that the Apple Watch, the most widely utilized wearable device, exhibits a sensitivity of only 25% for atrial flutter/atrial tachycardia (AFL/AT) (40), while the Fitbit demonstrates even lower accuracy in diagnosing atrial arrhythmias (41). The purpose of this study is to develop an ECG-AI model that accurately detects arrhythmia. Compared to traditional methods, this deep learning approach eliminates the need for manual design and selection of features, significantly reducing the expertise and time required for arrhythmia diagnosis. Moreover, by diminishing the reliance on specialized knowledge, the model can be deployed on everyday wearable devices such as the Apple Watch to facilitate routine monitoring of arrhythmias. This will assist clinicians in obtaining more precise ECG diagnoses and providing technical support for wearable and hospital devices.

In recent years, many studies have explored the application of deep learning models in arrhythmia classification, some of which focused on the problem of five categories of arrhythmias. Although these studies have demonstrated the potential of AI in arrhythmia detection, existing models vary significantly in complexity, accuracy, and computational efficiency. Our proposed CANet model significantly improves on traditional methods, especially in terms of lightweight design, real-time detection capabilities, and portability, which are critical for the clinical application of wearable devices. For example, Attia et al. (8) proposed an AI-assisted ECG algorithm for identifying atrial fibrillation (AF) in sinus rhythm with an accuracy of 83.3%.

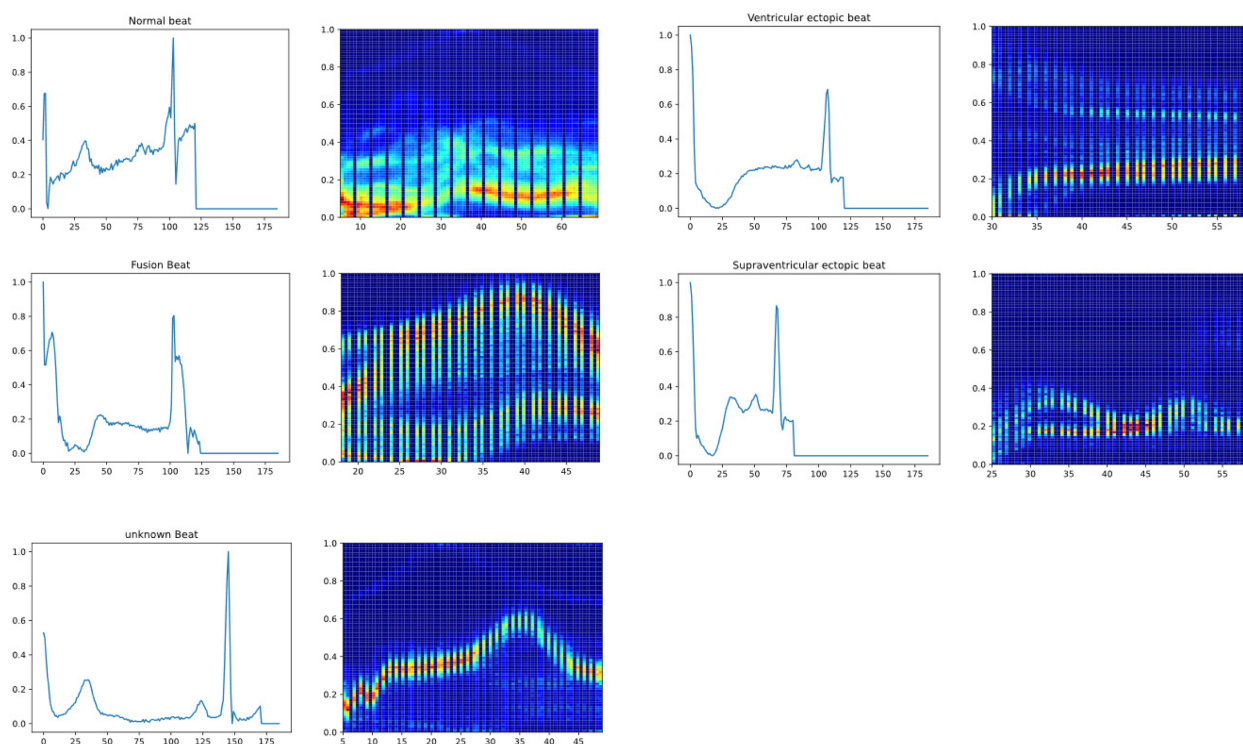


FIGURE 6

The heat activation map of new model (From left to right, up and down are respectively normal beat, ventricular ectopic beat, fusion beat, supraventricular ectopic beat, unknown beat).

However, the study only focused on one type of arrhythmia, while our model solves the problem of multi-class classification and can distinguish between normal, atrial, ventricular, atrial fibrillation, and unclassified arrhythmias, providing a more comprehensive solution. In addition, CANet integrates a multi-head attention mechanism that can simultaneously focus on multiple features of the ECG signal, improving accuracy without increasing computational complexity, which is also a limitation of many existing models. Similarly, Faour et al. (11) studied a convolutional neural network (CNN)-based AI model for detecting ST-segment elevation myocardial infarction. Although the model was effective in detecting specific types of arrhythmias, its high model complexity limited its application on portable devices. In contrast, CANet significantly reduced the number of parameters and improved computational efficiency by using deep separable convolutions, making it suitable for real-time arrhythmia detection on wearable ECG devices. Finally, the deep neural network (DNN) developed by Hannun et al. (7) showed high sensitivity and specificity in classifying arrhythmias from raw ECG data. Although this method performed strongly, its application in portable devices was limited by its model complexity and large size. Our model maintains high performance while adopting a lightweight design, enabling real-time detection and making it more suitable for deployment in resource-constrained environments. Compared with these studies, CANet not only achieves competitive accuracy, but also further improves performance without increasing computational overhead by integrating an efficient architecture and

attention mechanism to optimize feature extraction. This makes it particularly suitable for clinical environments that require real-time continuous monitoring and portability. In addition, CANet uses data augmentation techniques such as time warping and noise injection to enable it to perform well even with limited data, which is a scenario that many existing models have difficulty handling.

Of course, there are some limitations to this study. Although the model demonstrated superior accuracy compared to others during validation, it still presents a risk of misdiagnosis, particularly in the classification of ventricular ectopic beats and normal beats, where accuracy is lower. This suggests that the model's effectiveness in arrhythmia detection may require further enhancement through strategies such as data augmentation and balancing the dataset by oversampling rare classes or undersampling common ones to reduce bias towards specific patterns. As this study focuses on the development of deep learning methodologies, and given the limitations of dataset size and volume, comprehensive testing across a broad spectrum of cardiac arrhythmia classifications was not conducted. Consequently, the training and validation of this study have certain limitations. Given the current model's superior classification performance, we plan to incorporate additional data in future work and train the model to recognize a wider array of arrhythmia types, thus broadening its applicability. Additionally, the model's external testing process has certain limitations: while external validation typically requires testing across data from diverse sources, this model's external test was performed on a held-out set from the MIT-BIH dataset, which introduces some

bias and reduces the model's robustness. Furthermore, this study was not tested across extensive datasets, diverse populations, or complex clinical settings. Therefore, it is necessary to further evaluate the model's performance using datasets with broader coverage. Specifically, the current publicly available database used can only detect the types of arrhythmias mentioned earlier, and in the future, we plan to expand to include additional relevant types. Additionally, there is a lack of external test data in this study, and we may conduct a multi-center study to improve this in the future. Lastly, the system has not been developed or deployed yet, and further development is required for clinical translation.

5 Conclusions

In this endeavor, researchers have created an AI-powered ECG model. The model has superior performance in both five-fold cross-validation and external validation compared to other models, indicating its greater feasibility and potential for reliably detecting arrhythmias through a fusion deep learning model featuring separable convolution. The principle aim of the study is to facilitate the procurement of more precise ECG diagnoses through technological advancement and to augment technical support for wearable and clinical devices utilized in clinical circumstances.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

YH: Writing – original draft, Writing – review & editing, Conceptualization, Software, Methodology, Visualization, Project administration. YZ: Writing – original draft, Writing – review & editing, Visualization, Software, Methodology, Funding acquisition. YQ: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – original draft. JL: Data curation, Methodology, Software, Writing – original draft. JZ: Data curation, Investigation, Methodology, Writing – review & editing. DL: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Writing – review & editing.

QW: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the [Guizhou Provincial Health and Wellness Commission] under Grant [gzwkj 2021-102]; The National Natural Science Foundation of China (No. 82260084); and Guizhou Provincial Science and Technology Agency Project (Qian Ke He Foundation ZK [2023] General 217, Qian Ke He Foundation ZK [2022] General 268), and Support by Key Advantageous Discipline Construction Project of Guizhou Provincial Health Commission in 2023.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2024.1473482/full#supplementary-material>

SUPPLEMENTARY TABLE S1
Result of five-fold cross validation.

SUPPLEMENTARY TABLE S2
Result of external validation.

References

- Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* (2020) 396:1204–22. doi: 10.1016/S0140-6736(20)30925-9
- Wang Z, Chen Z, Wang X, Zhang L, Li S, Tian Y, et al. The disease burden of atrial fibrillation in China from a national cross-sectional survey. *Am J Cardiol*. (2018) 122:793–8. doi: 10.1016/j.amjcard.2018.05.015
- Deng J, Jiang Y, Chen ZB, Rhee J-W, Deng Y, Wang ZV. Mitochondrial dysfunction in cardiac arrhythmias. *Cells*. (2023) 12:679. doi: 10.3390/cells12050679
- Tracy CM, Epstein AE, Darbar D, Dimarco JP, Dunbar SB, Estes NM III, et al. 2012 ACCF/AHA/HRS focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American college of cardiology foundation/American heart association task force on practice guidelines and the heart rhythm society. *Circulation*. (2012) 126:1784–800. doi: 10.1161/CIR.0b013e3182618569

5. Sun T, Ye M, Lei F, Qin J-J, Liu Y-M, Chen Z, et al. Prevalence and trend of atrial fibrillation and its associated risk factors among the population from nationwide health check-up centers in China, 2012–2017. *Front Cardiovasc Med.* (2023) 10:1151575. doi: 10.3389/fcvm.2023.1151575
6. Begg G, Willan K, Tyndall K, Pepper C, Tayebjee MJB. Electrocardiogram interpretation and arrhythmia management: a primary and secondary care survey. *Br J Gen Pract.* (2016) 66:e291–6. doi: 10.3399/bjgp16X684781
7. Hannun AY, Rajpurkar P, Haghighian M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* (2019) 25(1):65–9. doi: 10.1038/s41591-018-0268-3
8. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, 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 (London, England).* (2019) 394(10201):861–7. doi: 10.1016/S0140-6736(19)31721-0
9. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation.* (2022) 145:122–33. doi: 10.1161/CIRCULATIONAHA.121.057480
10. Vázquez CG, Breuss A, Gnarr O, Portmann J, Madaffari A, Da Poian G. Label noise and self-learning label correction in cardiac abnormalities classification. *Physiol Meas.* (2022) 43:094001. doi: 10.1088/1361-6579/ac89cb
11. Faour A, Cherrett C, Gibbs O, Lintern K, Mussap CJ, Rajaratnam R, et al. Utility of prehospital electrocardiogram interpretation in ST-segment elevation myocardial infarction utilizing computer interpretation and transmission for interventional cardiologist consultation. *Catheter Cardiovasc Interv.* (2022) 100(3):295–303. doi: 10.1002/ccd.30300
12. Somani S, Russak AJ, Richter F, Zhao S, Vaid A, Chaudhry F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *Europace.* (2021) 23:1179–91. doi: 10.1093/europace/eaab377
13. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:e215–20. doi: 10.1161/01.cir.101.23.e215
14. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol Mag.* (2001) 20:45–50. doi: 10.1109/51.932724
15. Singh BN, Tiwari AK. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digit Signal Process.* (2006) 16:275–87. doi: 10.1016/j.dsp.2005.12.003
16. Mallat S. An efficient image representation for multiscale analysis. In: *Machine Vision*. Optica Publishing Group (1987). p. FD4. doi: 10.1364/MV.1987.FD4
17. Liu B, Li Y. ECG Signal denoising based on similar segments cooperative filtering. *Biomed Signal Process Control.* (2021) 68:102751. doi: 10.1016/j.bspc.2021.102751
18. Talmon J, Kors J, Van Bommel J. Adaptive Gaussian filtering in routine ECG/VCG analysis. *IEEE Trans Acoust Speech Signal Process.* (1986) 34:527–34. doi: 10.1109/TASSP.1986.1164864
19. Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng.* (1985) 32:230–6. doi: 10.1109/TBME.1985.325532
20. Sathyapriya L, Murali L, Manigandan T. Analysis and detection R-peak detection using modified pan-tompkins algorithm. 2014 *IEEE International Conference on Advanced Communications, Control and Computing Technologies: IEEE* (2014). p. 483–7
21. Musa N, Gital AYU, Aljojo N, Chiroma H, Adewole KS, Mojeed HA, et al. A systematic review and meta-data analysis on the applications of deep learning in electrocardiogram. *J Ambient Intell Humaniz Comput.* (2023) 14:9677–750. doi: 10.1007/s12652-022-03868-z
22. Salloum R, Kuo C-CJ. ECG-based biometrics using recurrent neural networks. 2017 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): IEEE* (2017). p. 2062–6
23. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *NeurIPS Proceedings* (2017). P. 30.
24. Fran C. Deep learning with depth wise separable convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
25. Baloglu UB, Talo M, Yildirim O, San Tan R, Acharya UR. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognit Lett.* (2019) 122:23–30. doi: 10.1016/j.patrec.2019.02.016
26. Abdullah RM, Abdulazeez AM. Electrocardiogram classification based on deep convolutional neural networks: a review. *Fusion Pract Appl.* (2021) 3:43–43–53. doi: 10.54216/FPA.030103
27. Prabhakararao E, Dandapat S. Attentive RNN-based network to fuse 12-lead ECG and clinical features for improved myocardial infarction diagnosis. *IEEE Signal Process Lett.* (2020) 27:2029–33. doi: 10.1109/LSP.2020.3036314
28. Noor ST, Asad ST, Khan MM, Gaba GS, Al-Amri JF, Masud M. Predicting the risk of depression based on ECG using RNN. *Comput Intell Neurosci.* (2021) 2021:1299870. doi: 10.1155/2021/1299870
29. Memory LS-T. Long short-term memory. *Neural Comput.* (2010) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735
30. Jyotishi D, Dandapat S. An LSTM-based model for person identification using ECG signal. *IEEE Sens Lett.* (2020) 4:1–4. doi: 10.1109/LSSENS.2020.3012653
31. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv.* (2014). doi: 10.48550/arXiv.1412.3555
32. Graves A, Schmidhuber J. Framework phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* (2005) 18:602–10. doi: 10.1016/j.neunet.2005.06.042
33. Xu X, Jeong S, Li J. Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM. *IEEE Access.* (2020) 8:125380–88. doi: 10.1109/ACCESS.2020.3006707
34. Cheng J, Zou Q, Zhao Y. ECG signal classification based on deep CNN and BiLSTM. *BMC Med Inform Decis Mak.* (2021) 21:1–12. doi: 10.1186/s12911-020-01362-0
35. Li J, Ke L, Du Q, Ding X, Chen X. Research on the classification of ECG and PCG signals based on BiLSTM-GoogLeNet-DS. *Appl Sci.* (2022) 12:11762. doi: 10.3390/app122211762
36. Holst H, Ohlsson M, Peterson C, Edenbrandt L. A confident decision support system for interpreting electrocardiograms. *Clin Physiol.* (1999) 19:410–8. doi: 10.1046/j.1365-2281.1999.00195.x
37. Report TWCOT, Health OC, China DI, Hu S-SJ. Report on cardiovascular health and diseases in China 2021: an updated summary. *J Geriatr Cardiol.* (2023) 20:399–430. doi: 10.26599/1671-5411.2023.06.001
38. Norcini JJ, Weng W, Boulet J, McDonald F, Lipner RS. Associations between initial American board of internal medicine certification and maintenance of certification status of attending physicians and in-hospital mortality of patients with acute myocardial infarction or congestive heart failure: a retrospective cohort study of hospitalisations in Pennsylvania, USA. *BMJ Open.* (2022) 12:e055558. doi: 10.1136/bmjopen-2021-055558
39. Pay L, Yumurtas AÇ, Satti DI, Hui JMH, Chan JSK, Mahalwar G, et al. Arrhythmias beyond atrial fibrillation detection using smartwatches: a systematic review. *Anatol J Cardiol.* (2023) 27:126. doi: 10.14744/AnatolJCardiol.2023.2799
40. Caillol T, Strik M, Ramirez FD, Abu-Alrub S, Marchand H, Buliard S, et al. Accuracy of a smartwatch-derived ECG for diagnosing bradyarrhythmias, tachyarrhythmias, and cardiac ischemia. *Circ Arrhythm Electrophysiol.* (2021) 14:e009260. doi: 10.1161/CIRCEP.120.009260
41. Koshy AN, Sajeev JK, Nerlekar N, Brown AJ, Rajakariar K, Zureik M, et al. Smart watches for heart rate assessment in atrial arrhythmias. *Int J Cardiol.* (2018) 266:124–7. doi: 10.1016/j.ijcard.2018.02.073



OPEN ACCESS

EDITED BY

Evangelos Oikonomou,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Panteleimon Pantelidis,
National and Kapodistrian University of Athens,
Greece

*CORRESPONDENCE

Arijita Banerjee

✉ b.arijita@gmail.com;

✉ arijita@abcmrc.iitkgp.ac.in

RECEIVED 19 December 2024

ACCEPTED 31 January 2025

PUBLISHED 14 February 2025

CITATION

Banerjee A (2025) Artificial intelligence enabled mobile health technologies in arrhythmias-an opinion article on recent findings.

Front. Cardiovasc. Med. 12:1548554.

doi: 10.3389/fcvm.2025.1548554

COPYRIGHT

© 2025 Banerjee. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence enabled mobile health technologies in arrhythmias-an opinion article on recent findings

Arijita Banerjee*

Department of Physiology, IIT Kharagpur, Kharagpur, India

KEYWORDS

artificial intelligence (AI), machine learning (ML), atrial fibrillation, supraventricular tachycardia (SVT), mobile health (mHealth), ventricular arrhythmia

Introduction

Integration of ion channels and transporters and inherent property of automaticity of myocardial cells are necessary for the transmission of electrical impulses throughout the myocardium and the generation of a normal cardiac rhythm. When either of these, normal electrophysiological process of impulse generation or normal conduction of action potential gets disrupted, patients experience cardiac arrhythmias. The risk of acquired arrhythmias is significantly increased in presence of structural heart diseases, myocardial infarction and metabolic disorders. The majority of cardiac arrhythmias are categorised according to the rate at which they generate impulses or by where they originate in the myocardium. These include atrial fibrillation (AF), atrial flutter, ventricular tachycardia (VT), supraventricular tachycardia (SVT), ventricular fibrillation and bradyarrhythmias (1). With its rapid and erratic electrical signals in the atria, AF is the most prevalent type and causes ineffective contractions. AF patients present with shortness of breath, exhaustion, palpitations, and a higher risk of stroke. Anticoagulation therapy to prevent thromboembolic events along with antiarrhythmic medications, are common management strategies. Sudden cardiac arrest caused by ventricular arrhythmias results in patients losing consciousness. In these situations, immediate cardiopulmonary resuscitation (CPR) and defibrillation are critical for survival (2). Global estimates indicate that cardiac arrhythmias impact nearly 2% of the world's population and are linked to significant socioeconomic burden. According to recent research, machine learning algorithms may enhance the risk stratification for long-term cardiac arrhythmia. The development of mobile health technologies has provided customer-focused health care opportunities (3). In this opinion, the potential applications of the current and upcoming mHealth technologies for treating cardiac arrhythmias are illustrated.

Role of digital technologies in arrhythmia care

Contemporary portable gadgets designed for health monitoring, such as photoplethysmography and ECG systems (4) are not only affordable but the high-speed internet access of these sensors have enabled patients to access healthcare more widely. These technologies enable real-time monitoring and early arrhythmia detection, allowing patients to better manage their conditions and receive timely medical intervention (5). Using longer-term event recorders, 24–48 Holter, or medically certified ambulant

electrocardiogram (ECG) monitors has been necessary for the crucial step of correlating symptoms with rhythm. Long-term rhythm monitoring is made possible by mHealth devices, especially smartphone-based ECG and PPG technology, which is reasonably priced (6). Cardiac electrophysiology has been profoundly modified by AI and machine learning, with advancements in mobile technology enabling the measurement of heart-related physiological data. The healthcare industry now has access to a wealth of data, including accelerometers, ECGs, and PPG signals (7).

Atrial fibrillation screening

New developments in current contact-free plethysmography using smartphone cameras on the face and fingers have demonstrated promise for examining atrial fibrillation. AI algorithms demonstrate a high degree of cardiac rhythm discrimination when used for ECG readings, including those recorded with mobile cardiac telemetry. It will take thorough algorithm validation, data integration with the healthcare system, improvement of current clinical workflows, and strong patient access to turn these fascinating findings that lead to better clinical results expanded to unprecedented levels. Some excellent examples of recent innovations include the TeleCheck-AF

project, smartphone ECG surveillance, and home antiarrhythmic medication loading with smartphone tracings (8).

Smartwatches can detect irregular pulses, confirming AF diagnosis through ECG patch monitoring, as demonstrated in studies like Apple Heart Study (9) and Fitbit and Huawei Heart Studies (10). The HEARTLINE trial explores the impact of accessible devices like Apple Watch on early AF detection and clinical outcomes, while the LOOP trial explores anticoagulant use (11, 12). Population-wide AF screening may benefit, but its usability and detection depend on the screening modality and population characteristics. Diagnosis of AF is typically made through ECG or continuous ECG recording, but concerns about false-positive diagnoses arise from sensitivity and specificity variations. The detection rates of new AF have only ranged from 0.9%–7.4% using handheld ECG monitors like Merlin, which rely on automated algorithms and have sensitivities ranging from 93%–100% (13). PPG and ECG-based wearables can identify AF in patients with previous AF, anti-arrhythmic medications, cardioversion, or ablation as highlighted in Table 1. While ECG is still the gold standard for diagnosing AF, PPG is already found in the majority of commercially available smartphones and wearable technology, making it a low-cost way to monitor arrhythmias even if it doesn't allow for precise diagnosis. Further ECG evaluation is necessary to confirm arrhythmias identified by PPG alone.

TABLE 1 Validated devices involving mHealth technologies for arrhythmia.

Type of Device	Study	Sensitivity (%)	Specificity (%)	Arrhythmia	Pros	Cons
Hand-held ECG devices	Zenico (14)	96	92	AF	Cloud-based analysis service	No display of ECG tracings
	MyDiagnostick (15)	100	96	AF	Recording and storage device	Rechargeable battery not available
	Merlin ECG event recorders (13)	93.9	90.1	AF	Display of ECG recordings	Rechargeable battery not available
Wearable PPG based	Apple Watch (9)	98	90.2	AF/flutter/SVT	Simple technology. Easy to carry. FDA approved	Blood pressure cannot be measured. PPG sensor consumes more power. Expensive.
	Fitbit (10)	68	98	AF/flutter/SVT	Easy to carry. FDA approved	Blood pressure cannot be measured. Expensive.
	AliveCor Kardia (16)	98	97	AF/flutter	ECG recordings displayed	iphone/android is required
Smartphone ECG based	AliveCor (17)	Alivecor is accurate in measuring QTc interval ($P < .01$)		QTc in sinus rhythm Easy to carry	ECG recordings displayed	iphone/android is required
	AliveCor Kardia (18)	89	91	SVT	ECG recordings displayed Easy to carry FDA approved	iphone/android is required
	Pulse-smart (36)	97	93.5	AF, PVC	Can distinguish between sinus rhythm and irregular abnormal pulse. Easy to use.	Noise and light affect accuracy.
Smartphone PPG based	Cardiio Rhythm (19)	92.9	97.7	AF/flutter	Low cost. Automatic beat to beat measurement of HR.	Android application needed
Wearable ECG based	Necklace-ECG (20)	99.1	98.5	AF	ECG tracings available. Easy, simple.	Continuous wearing is required.
	Samsung Simband 2.0 (21)	98.2	98.1	AF/flutter	Both ECG and PPG recordings are obtained.	Android is necessary
Holter based-continuous monitoring	Zio Patch (22)	96 arrhythmia events detected		AF, SVT, VT	No battery charging required FDA approved	Prolonged wearing

ECG, electrocardiography; PPG, photoplethysmography; AF, atrial fibrillation; PVC, premature ventricular contraction; SVT, supraventricular tachycardia; VT, ventricular tachycardia; QTc, corrected QT interval; HR, heart rate.

Atrial fibrillation management and the potential risk of stroke

The potential of PPG and ECG-based wearables to identify AF in patients with a history of AF, anti-arrhythmic medication use, cardioversion, and ablation has been confirmed by certain studies using artificial neural network (23). Integrating mobile technologies with a “pill in the pocket” approach could provide benefits like closer monitoring, early antiarrhythmic medication administration, confirmatory validation of symptoms, and avoiding side effects and emergency medical visits (24).

In order to predict paroxysmal AF on ECGs from patients in sinus rhythm, the Mayo Clinic created the first AI-ECG algorithm using almost 650,000 ECGs. Furthermore, the use of AI-ECG for AF estimation was investigated in patients who had an embolic stroke of unknown cause, in which silent underlying AF is often suspected to be the cause (25). AI/ML techniques may also present the chance to stratify patients based on outcomes, like the likelihood of a stroke or the success of cardioversion, in the context of a new AF diagnosis. The ML models outperformed the CHA2DS2-VASc and HATCH scores in predicting the risk for ischemic stroke, but they were less effective than the scores in predicting 6-month AF recurrence, 6-month rhythm control, and pharmacological cardioversion success (26, 27). Regardless of the AF pattern (silent or paroxysmal) or whether the AF burden is low due to automatic cessation of rhythm control techniques, current guidelines advise lifelong anticoagulation based on risk factors.

Role of mHealth in ventricular and supraventricular arrhythmias

SVT is challenging to diagnose due to its unpredictable nature and lack of diagnostic yields. Traditional methods have diagnostic yields ranging from 10% to 50%–60%. mHealth devices offer long-term, affordable rhythm monitoring, effectively diagnosing patients experiencing brief episodes of prolonged palpitations. Smartphone-based single-lead ECGs have a high resolution to distinguish SVT from sinus tachycardia misdiagnosis (89% sensitivity and 91% specificity). However, only 51% of surveyed doctors would proceed with an invasive EP study based on symptomatic, regular tachycardia (18). mHealth devices may mistakenly diagnose palpitations caused by PVCs as AF due to irregular rhythms. Discrimination algorithms could address this issue. Smartphone-based algorithms have successfully distinguished PVCs from sinus rhythm, PACs, and AF with 96% accuracy. A computational algorithm created a feature matrix from QRS attributes from a smartphone-connected ECG device, showing 98.69% PVC recognition accuracy. Smartphones may be helpful in diagnosing ventricular arrhythmias, as evidenced by case reports (28, 29). According to the 2019 guidelines published by the European Society of Cardiology (ESC), mobile recording devices may be required for the diagnosis of supraventricular tachycardias (SVT) due to their ease of use, but validation is necessary. EP-guided ablation, a potential treatment, can be

accelerated with the use of smartphone-based one-lead ECG recordings. Unfortunately, quantifying burden is difficult due to the irregular nature of SVTs (30).

Use of mHealth application for arrhythmia care in children

Paediatric and congenital heart populations are quickly adopting mHealth technologies, despite the fact that these tools were created and validated in the adult population. When compared to traditional 12-lead ECGs in children, a few mHealth devices, like the Apple Watch and Alive Cor Kardia Monitor (31), have been evaluated for symptom-rhythm correlation and QT evaluation with high quality data (32). ICDs are effective in saving lives for patients with high risk of sudden cardiac death (SCD), but they don't significantly reduce sudden deaths. Machine learning can develop algorithms to recognize reduced left ventricular function from a 12-lead ECG, which predicts ICD benefit (33). Traditional markers and AI-based markers struggle to improve mortality rates by predicting positive and negative values and identifying modifiable physiological processes. Wearable technology measuring heart rate variability (HRV) can improve general health, but its efficacy is limited due to limited data in controlled settings (34).

Limitations of mHealth technologies

For various stakeholders, integrating digital health technologies into the treatment of patients with arrhythmias poses a number of challenges. Healthcare providers encounter challenges like a lack of knowledge about the features of the devices, a lack of confidence in their use, and worries about liability. Challenges for patients and consumers include the need for additional testing, socioeconomic disparities that impact access, potential anxiety related to test results, and differing levels of digital literacy. Operational difficulties include continuous charging of devices, cybersecurity threats, data storage problems, insufficient investment in workflows to handle the growing number of devices, and the incorporation of device data into electronic health records. Additionally, the adoption of these technologies in the healthcare landscape is made more difficult by the absence of clear guidance on legal obligations and reimbursement proceedings (35).

Conclusion

High cardiac rhythm discrimination is demonstrated by various AI algorithms in ECG readings. According to recent research, AI algorithms may enhance the risk stratification for long-term ventricular arrhythmia. However, for widespread clinical results, additional validation, data integration, clinical workflow enhancements, and patient access are required. Data from wearable devices being incorporated into electronic medical records (EMRs) is crucial for efficient clinical review and decision-making. However, clinician time is a significant barrier to this integration.

By reducing the need for frequent office visits and intervention, these systems can ensure consistent and effective patient care. Clinical data is essential for confirming device accuracy and determining the effectiveness of interventions based on findings. Progressive automation may be an option, but systems should begin as semi-automated. However, to evaluate the clinical usefulness of machine learning models in enhancing subsequent ventricular rhythm disturbances, additional investigations with bigger sample sizes, robust validity, more varied patient samples, are required.

Author contributions

AB: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

References

1. Tse G. Mechanisms of cardiac arrhythmias. *J Arrhythmia*. (2016) 32:75–81. doi: 10.1016/j.joa.2015.11.003
2. Wang M, Tu X. The genetics and epigenetics of ventricular arrhythmias in patients without structural heart disease. *Front Cardiovasc Med*. (2022) 9:891399. doi: 10.3389/fcvm.2022.891399
3. Amuthan R, Curtis AB. Sex-specific considerations in drug and device therapy of cardiac arrhythmias: JACC focus seminar 6/7. *J Am Coll Cardiol*. (2022) 79:1519–29. doi: 10.1016/j.jacc.2021.11.066
4. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable devices for ambulatory cardiac monitoring: JACCstate-of-the-art review. *J Am Coll Cardiol*. (2020) 75:1582–92. doi: 10.1016/j.jacc.2020.01.046
5. Sanders D, Ungar L, Eskander MA, Seto AH. Ambulatory ECG monitoring in the age of smartphones. *Cleve Clin J Med*. (2019) 86:483–93. doi: 10.3949/ccjm.86a.18123
6. Shabaan M, Arshid K, Yaqub M, Jinchao F, Zia MS, Bojja GR, et al. Survey: smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis. *BMC Med Inf Decis Mak*. (2020) 20:177. doi: 10.1186/s12911-020-01199-7
7. Yan BP, Lai WHS, Chan CKY, Au ACK, Freedman B, Poh YC, et al. High-throughput, contact-free detection of atrial fibrillation from video with deep learning. *JAMA Cardiol*. (2020) 5:105–7. doi: 10.1001/jamacardio.2019.4004
8. Gawalko M, Duncker D, Manninger M, Velden RMJ, Hermans ANL, Verhaert DVM, et al. The European TeleCheck-AF project on remote app-based management of atrial fibrillation during the COVID-19 pandemic: centre and patient experiences. *Europace*. (2021) 23:1003–15. doi: 10.1093/europace/euab050
9. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. *J Am Coll Cardiol*. (2019) 74:2365–75. doi: 10.1016/j.jacc.2019.08.019
10. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the fitbit heart study. *Circulation*. (2022) 146:1415–24. doi: 10.1161/CIRCULATIONAHA.122.060291
11. Rizas KD, Freyer L, Sappler N, von Stülpnagel L, Spielbichler P, Krasniqi A, et al. Smartphone-based screening for atrial fibrillation: a pragmatic randomized clinical trial. *Nat Med*. (2022) 28:1823–30. doi: 10.1038/s41591-022-01979-w
12. Svendsen JH, Diederichsen SZ, Hojberg S, Krieger DW, Graff C, Kronborg C, et al. Implantable loop recorder detection of atrial fibrillation to prevent stroke (the LOOP study): a randomised controlled trial. *Lancet*. (2021) 398:1507–16. doi: 10.1016/S0140-6736(21)01698-6
13. Kearley K, Selwood M, Van den Bruel A, Thompson M, Mant D, Hobbs FR, et al. Triage tests for identifying atrial fibrillation in primary care: a diagnostic accuracy study comparing single-lead ECG and modified BP monitors. *BMJ Open*. (2014) 4(5):e004565. doi: 10.1136/bmjopen-2013-004565
14. Doliwa PS, Frykman V, Rosenqvist M. Short-term ECG for out of hospital detection of silent atrial fibrillation episodes. *Scand Cardiovasc J*. (2009) 43(3):163–8. doi: 10.1080/14017430802593435
15. Tieleman RG, Plantinga Y, Rinkes D, Bartels GL, Posma JL, Cator R, et al. Validation and clinical use of a novel diagnostic device for screening of atrial fibrillation. *Europace*. (2014) 16(9):1291–5. doi: 10.1093/europace/euu057
16. Lau JK, Lowres N, Neubeck L, Brieger DB, Sy RW, Galloway CD, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. *Int J Cardiol*. (2013) 165(1):193–4. doi: 10.1016/j.ijcard.2013.01.220
17. Garabelli P, Stavrakis S, Albert M, Koomson E, Parwani P, Chohan J, et al. Comparison of QT interval readings in normal sinus rhythm between a smartphone heart monitor and a 12-lead ECG for healthy volunteers and inpatients receiving sotalol or dofetilide. *J Cardiovasc Electrophysiol*. (2016) 27:827–32. doi: 10.1111/jce.12976
18. Wegner FK, Kochhäuser S, Frommeyer G, Lange PS, Ellermann C, Leitz P, et al. Prospective blinded evaluation of smartphone based ECG for differentiation of supraventricular tachycardia from inappropriate sinus tachycardia. *Clin Res Cardiol*. (2021) 110:905–12. doi: 10.1007/s00392-021-01856-5
19. Chan PH, Wong CK, Poh YC, Pun L, Leung WW, Wong YF, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. *J Am Heart Assoc*. (2016) 5(7):e003428. doi: 10.1161/JAHA.116.003428
20. Santala OE, Lipponen JA, Jäntti H, Rissanen TT, Halonen J, Kolk I, et al. Necklace-embedded electrocardiogram for the detection and diagnosis of atrial fibrillation. *Clin Cardiol*. (2021) 44:620–6. doi: 10.1002/clc.23580
21. Ding EY, Han D, Whitcomb C, Bashar SK, Adaramola O, Soni A, et al. Accuracy and usability of a novel algorithm for detection of irregular pulse using a smartwatch among older adults: observational study. *JMIR Cardio*. (2019) 3(1):e13850. doi: 10.2196/13850
22. Barrett PM, Komatireddy R, Haaser S, Topol S, Sheard J, Encinas J, et al. Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. *Am J Med*. (2014) 127(95):95.e11–17. doi: 10.1016/j.amjmed.2013.10.003

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

23. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* (2019) 25(1):65–9. doi: 10.1038/s41591-018-0268-3
24. Lyth J, Svennberg E, Bernfort L, Aronsson M, Frykman V, Al-Khalili F, et al. Cost-effectiveness of population screening for atrial fibrillation: the STROKESTOP study. *Eur Heart J.* (2023) 44:196–204. doi: 10.1093/eurheartj/ehac547
25. Attia ZI, Noseworthy P, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, 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) 394:861–7. doi: 10.1016/S0140-6736(19)31721-0
26. Buck BH, Hill MD, Quinn FR, Butcher KS, Menon BK, Gulamhusein S, et al. Effect of implantable vs prolonged external electrocardiographic monitoring on atrial fibrillation detection in patients with ischemic stroke: the PER DIEM randomized clinical trial. *JAMA.* (2021) 325:2160–8. doi: 10.1001/jama.2021.6128
27. Bernstein RA, Kamel H, Granger CB, Piccini JP, Sethi PP, Katz JM, et al. Effect of long-term continuous cardiac monitoring vs usual care on detection of atrial fibrillation in patients with stroke attributed to large- or small-vessel disease: the STROKE-AF randomized clinical trial. *JAMA.* (2021) 325:2169–77. doi: 10.1001/jama.2021.6470
28. Kumar S, Banerjee A. Artificial intelligence-enabled smartwatch used for the detection of idiopathic ventricular tachycardia: a case report. *Cureus.* (2023) 15(7):e42054. doi: 10.7759/cureus.42054
29. Ringwald M, Crich A, Beysard N. Smart watch recording of ventricular tachycardia: case study. *Am J Emerg Med.* (2020) 38:849.e3–e845. doi: 10.1016/j.ajem.2019.10.040
30. Manninger M, Zweiker D, Svennberg E, Chatzikyriakou S, Pavlovic N, Zaman JAB, et al. Current perspectives on wearable rhythm recordings for clinical decision-making: the wEHRables 2 survey. *EP Europace.* (2021) 23:1106–13. doi: 10.1093/europace/euab064
31. Kobel M, Kalden P, Michaelis A, Markel F, Mensch S, Weidenbach M, et al. Accuracy of the apple watch iECG in children with and without congenital heart disease. *Pediatr Cardiol.* (2022) 43:191–6. doi: 10.1007/s00246-021-02715-w
32. Karacan M, Celik N, Gul EE, Akdeniz C. Validation of a smartphone-based electrocardiography in the screening of QT intervals in children. *North Clin Istanbul.* (2018) 6:48–52. doi: 10.14744/nci.2018.44452
33. Kitsiou S, Vatani H, Paré G, Gerber BS, Buchholz SW, Kansal MM, et al. Effectiveness of mobile health technology interventions for patients with heart failure: systematic review and meta-analysis. *Can J Cardiol.* (2021) 37:1248–59. doi: 10.1016/j.cjca.2021.02.015
34. Singh N, Moneghetti KJ, Christle JW, Hadley D, Froelicher V, Plews D. Heart rate variability: an old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. Part two: prognosis and training. *Arrhythm Electrophysiol Rev.* (2018) 7:247–55. doi: 10.15420/aer.2018.30.2
35. Orchard JJ, Neubeck L, Orchard JW, Puranik R, Raju H, Freedman B, et al. ECG-based cardiac screening programs: legal, ethical, and logistical considerations. *Heart Rhythm.* (2019) 16(10):1584–91. doi: 10.1016/j.hrthm.2019.03.025
36. McMANUS DD, Chong JW, Soni A, Saczynski JS, Esa N, Napolitano C, et al. PULSE-SMART: pulse-based arrhythmia discrimination using a novel smartphone application. *J Cardiovasc Electrophysiol.* (2016) 27(1):51–7. doi: 10.1111/jce.12842



OPEN ACCESS

EDITED BY

Evangelos Oikonomou,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Weichieh Lee,
Chi Mei Medical Center, Taiwan
Mert İlker Hayıroğlu,
Dr. Siyami Ersek Chest Cardiovascular Surgery
Training and Research Hospital, Türkiye

*CORRESPONDENCE

Guoxiong Deng
✉ dengguoxiong@stu.gxmu.edu.cn

RECEIVED 02 November 2024

ACCEPTED 12 February 2025

PUBLISHED 26 February 2025

CITATION

Teng J and Deng G (2025) Bioinformatics analysis of the expression of potential common genes and immune-related genes between atrial fibrillation and chronic kidney disease.
Front. Cardiovasc. Med. 12:1521722.
doi: 10.3389/fcvm.2025.1521722

COPYRIGHT

© 2025 Teng and Deng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Bioinformatics analysis of the expression of potential common genes and immune-related genes between atrial fibrillation and chronic kidney disease

Jieying Teng^{1,2} and Guoxiong Deng^{1,2*}

¹Department of Cardiology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, China,
²Department of Cardiology, The First People's Hospital of Nanning, Nanning, China

Research objective: This study is based on bioinformatics analysis to explore the co-expressed differentially expressed genes (DEGs) between atrial fibrillation (AF) and chronic kidney disease (CKD), identify the biomarkers for the occurrence and development of the two diseases, investigate the potential connections between AF and CKD, and explore the associations with immune cells.

Methods: We downloaded Two AF gene chip datasets (GSE79768, GSE14975) and two CKD gene chip datasets (GSE37171, GSE120683) from the GEO database. After pre-processing and standardizing the datasets, two DEGs datasets were obtained. The DEGs were screened using R language, and the intersection was taken through Venn diagrams to obtain the co-expressed DEGs of AF and CKD. To obtain the signal pathways where the co-expressed DEGs were significantly enriched, GO/KEGG enrichment analyses were used to analysis the co-expressed DEGs. The Cytoscape software was used to further construct a PPI network and screen key characteristic genes, and the top 15 co-expressed DEGs were screened through the topological algorithm MCC. To further screen key characteristic genes, two machine-learning algorithms, LASSO regression and RF algorithm, were performed to screen key characteristic genes for the two disease datasets respectively to determine the diagnostic values of the characteristic genes in the two diseases. The GeneMANIA online database and NetworkAnalyst platform were used to construct gene-gene and TFs-gene interaction network diagrams respectively to predict gene functions and find key transcription factors. Finally, the correlation between key genes and immune cell subtypes was performed by Spearman analysis.

Research results: A total of 425 DEGs were screened out from the AF dataset, and 4,128 DEGs were screened out from the CKD dataset. After taking the intersection of the two, 82 co-expressed DEGs were obtained. The results of GO enrichment analysis of DEGs showed that the genes were mainly enriched in biological processes such as secretory granule lumen, blood microparticles, complement binding, and antigen binding. KEGG functional enrichment analysis indicated that the genes were mainly enriched in pathways such as the complement coagulation cascade, systemic lupus erythematosus, and Staphylococcus aureus infection. The top 15 DEGs were obtained through the MCC topological algorithm of Cytoscape software. Subsequently, based on LASSO regression and RF algorithm, the key characteristic genes of the 15 co-expressed DEGs of AF and CKD were further screened, and by taking the intersection through Venn diagrams, five key characteristic genes were finally obtained: PPBP, CXCL1, LRRK2, RGS18, RSAD2. ROC curves were constructed

to calculate the area under the curve to verify the diagnostic efficacy of the key characteristic genes for diseases. The results showed that RSAD2 had the highest diagnostic value for AF, and the diagnostic values of PPBP, CXCL1, and RSAD2 for CKD were all at a relatively strong verification level. Based on AUC >0.7, co-expressed key genes with strong diagnostic efficacy were obtained: PPBP, CXCL1, RSAD2. The results of the GeneMANIA online database showed that the two biomarkers, BBPB and CXCL1, mainly had functional interactions with cytokine activity, chemokine receptor activity, cell response to chemokines, neutrophil migration, response to chemokines, granulocyte chemotaxis, and granulocyte migration. The TFs-gene regulatory network identified FOXC1, FOXL1, and GATA2 as the main transcription factors of the key characteristic genes. Finally, through immune infiltration analysis, the results indicated that there were various immune cell infiltrations in the development processes of AF and CKD.

Research conclusion: PPBP, CXCL1, and RSAD2 are key genes closely related to the occurrence and development processes between AF and CKD. Among them, the CXCLs/CXCR signaling pathway play a crucial role in the development processes of the two diseases likely. In addition, FOXC1, FOXL1, and GATA2 may be potential therapeutic targets for AF combined with CKD, and the development of the diseases is closely related to immune cell infiltration.

KEYWORDS

atrial fibrillation, chronic kidney disease, bioinformatics, inflammatory, immune process, CXCLs/CXCR

1 Research background

Atrial fibrillation (AF) is the most common arrhythmia in clinical practice. In the past two decades, with the increase in exposure to risk factors, the aging of the population, and the prolonged survival period of patients with chronic diseases (1), AF has brought a heavy disease burden to society and medical services. AF not only affects the quality of life of patients but also increases the risk of complications such as heart failure, stroke, death, and dementia (2, 3). The etiology of AF is complex and can involve multiple systemic diseases. Various factors can increase the susceptibility to AF, including primary diseases cardiovascular diseases such as coronary heart disease, hypertension, valvular heart disease, cardiomyopathy, etc., endocrine diseases such as hyperthyroidism, respiratory diseases such as sleep apnea syndrome, and oxidative stress and other factors. Koray Kalenderoglu et al. (4) pointed out that in the development of AF, the inflammatory response is closely related to the onset and maintenance of atrial fibrillation. In addition, this study evaluated the levels of inflammatory markers after balloon ablation to predict the risk of AF recurrence. It was found that the inflammatory indexes, namely the uric acid/albumin ratio (UAR), the systemic immune–inflammation index (SII), and the CRP/albumin ratio (CAR), were independently associated with the risk of AF recurrence after balloon ablation. Moreover, it was discovered that CAR might be better at predicting the recurrence of AF.

Chronic kidney disease (CKD) is a chronic disease with abnormal kidney structure or function lasting for more than 3 months. The key points of its common diagnostic criteria are as follows (5): the appearance of albuminuria (UAER $\geq$ 30 mg/24 h or UACR $\geq$ 30 mg/g), abnormal urinary sediment, related tubular

lesions, histological abnormalities, structurally abnormal findings in imaging, or a history of kidney transplantation, etc. CKD is a disease that seriously threatens human health, apart from cardiovascular and cerebrovascular diseases, diabetes, malignant tumors, and other diseases. In recent years, the prevalence of CKD has been on the rise globally, and the prevalence rate in the general population has reached as high as 14.3% (6, 7). In the development process of chronic kidney disease, inflammatory responses play an indispensable role (8). Some studies have pointed out that in the early stage of kidney injury, neutrophils, the main recruited inflammatory cells, promote the production of reactive oxygen species (ROS) and trigger an inflammatory cascade reaction (9). The inflammatory response, as an important immune defense mechanism of the body, is a dynamic process of injury, anti-injury, and repair, and can lead to the occurrence of various chronic diseases.

It was first discovered by Robert Bright and others in 1,836 that there is an association between the physiology and pathology of the heart and the kidney, and it was described as “the delicate and highly interdependent relationship between the heart and the kidney”. With the continuous advancement of research on the heart and the kidney, a new concept- “Cardiorenal Syndrome (CRS)” (10, 11) has been proposed, which refers to a syndrome in which acute or chronic functional abnormalities occur in one of the heart and kidney organs, resulting in acute or chronic functional abnormalities in the other organ. The heart and kidney are both important organs that affect hemodynamics, and at the same time, they are both affected by hemodynamics, and there is a strong interaction between them (12). The pathogenesis of cardiorenal syndrome involves multiple aspects, including changes in hemodynamics, enhanced activity of the renin-angiotensin-aldosterone system (RAAS) (13), oxidative

stress, production of advanced glycation end-products, endoplasmic reticulum stress, and persistent chronic inflammation, etc., and its exact mechanism has not been fully elucidated. At present, regarding the pathophysiological mechanisms of AF and CKD, many studies have identified that chronic inflammatory responses play important roles in the occurrence and development of these two diseases. This study's purpose is to analysis the potential biological significance of biomarkers in AF and CKD, further investigate the immune characteristics of immune cells in these two diseases, and search for potential diagnostic markers and therapeutic targets.

2 Materials and methods

2.1 Acquisition of datasets

The gene expression profile datasets of AF and CKD were downloaded from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). Among them, GSE79768 and GSE14975 are AF-related datasets, with a total of 19 atrial fibrillation case groups and 17 sinus rhythm control groups. GSE37171 and GSE120683 are CKD-related datasets, with a total of 78 chronic kidney disease patients and 43 normal control groups. Among them, the AF dataset GSE14975 and the CKD dataset GSE120683 are used as external validation sets, and all of the above datasets are obtained from the GPL570 platform. The specific contents of the gene chip data of the two diseases are shown in **Table 1**, and the research analysis process is shown in **Figure 1**.

2.2 Preliminary data processing and DEGs screening

The downloaded microarray datasets are read in based on the R (4.4.1) software. The expression data of the selected samples are annotated according to the GPL570 platform file, and the probe ID conversion is carried out to transform the probe matrix into a gene matrix. When a probe corresponds to multiple gene names, the first gene in the front is retained through the delimiter. The data are corrected by using the Normalize Between Arrays function of the “limma” package. The box-plots of the samples after standardizing the original data are shown in **Figures 2A,B**, respectively; the PCA plots after performing principal component analysis are shown in **Figures 2C,D**. Obtain the grouping information of the datasets of the two diseases. Use the “limma” package in R software to conduct differential

expression analysis on the gene expression profile matrices of AF and CKD respectively. Sequentially, through constructing a comparison matrix, fitting a linear model, and performing Bayesian testing, all the differentially expressed genes (DEGs) of the two diseases are finally obtained. To further screen the DEGs, for both the gene data of AF and CKD, $P < 0.05$ and $|\log_2(FC)| > 0.585$ are selected as the screening thresholds. With the help of the “pheatmap” package, heatmaps and volcano plots are drawn for the two diseases based on the obtained data. Furthermore, use the “VennDiagram” package in R software to draw a Venn diagram to identify the overlapping part of the two sets of DEGs, and obtain the co-expressed differentially expressed genes of the two groups of disease data. This overlapping part of co-expressed DEGs is used for subsequent research and analysis.

2.3 Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses

Utilize the “org.Hs.eg.db” R package for human genome annotation to transform the gene names in the co-expressed DEGs data of the two diseases, which is obtained from the above mentioned differential analysis, into ENTREZ IDs. By means of the “ClusterProfiler” package in R software, conduct GO and KEGG enrichment analyses on the co-expressed DEGs. Meanwhile, the enrichment analyses need to meet the conditions: $p.adjust < 0.05$ and $q.value < 0.05$. The results of gene function enrichment analyses are visualized through the “ggplot2”, “pathview” and “enrichplot” R packages. Finally, using the “cnetplot” function of the “BiocManager” R package to screen out the Top4 results in GO analysis and the Top8 results in KEGG analysis, and visualizing the results.

2.4 Construction of protein–protein interaction (PPI) network

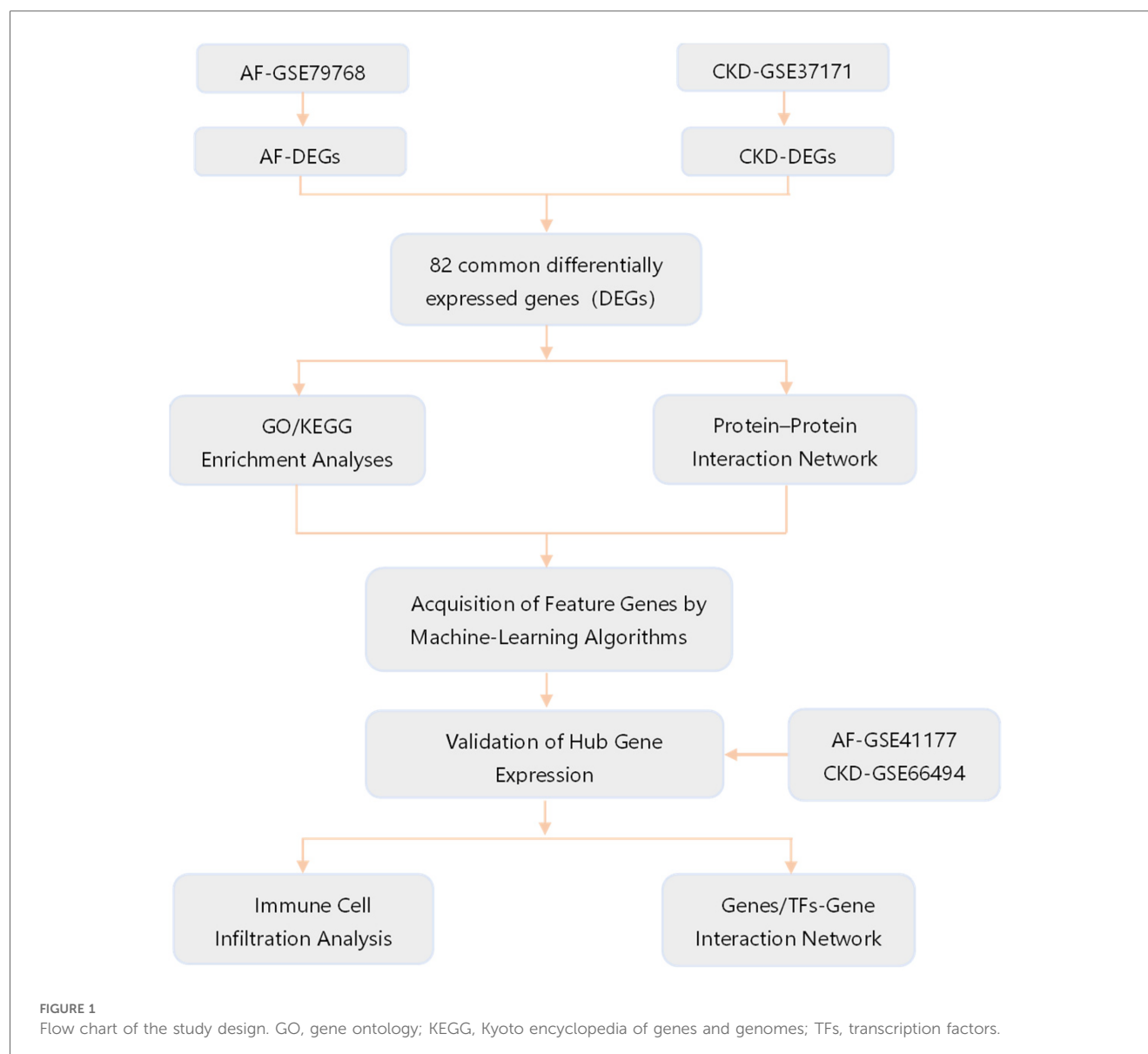
The co-expressed differentially expressed genes obtained after taking the intersection are imported into the STRING (<https://cn.string-db.org>) website, and the confidence score threshold is set to 0.4 to evaluate the interactions between genes by constructing a protein-interaction network diagram, and the screened results are downloaded. Then, with the help of Cytoscape software, the results obtained from the STRING database are visualized, and the top 15 genes are screened based on the topological algorithm-Matthews correlation coefficient (MCC).

2.5 Acquisition of feature genes by machine-learning algorithms

The top 15 co-expressed genes obtained by the MCC algorithm are further screened for key feature genes in AF and CKD respectively through two machine-learning algorithms. The least absolute shrinkage and selection operator (LASSO) and random

TABLE 1 Basic information of GEO datasets used in the study.

GSE series	Platform	Samples size		Disease
		Control	Disease	
GSE79768	GPL570	12	14	AF
GSE37171		40	75	CKD
GSE14975		5	5	AF
GSE120683		3	3	CKD



forest (RF) algorithms are used to obtain the key feature genes of the two diseases respectively, and the key feature genes of the two diseases are obtained by taking the intersection through Venn diagrams.

2.6 Verification of gene diagnostic efficacy with external datasets

The diagnostic values of the key characteristic genes obtained by machine-learning algorithms in the two diseases are evaluated by using the area under the ROC curve (AUC). The value of AUC ranges from 0.5 to 1. The closer it is to 1, the better the diagnostic performance and the higher the accuracy of the gene. The AF dataset GSE14975 (AF=5, SR=5) and the CKD dataset GSE120683 (CKD=3, Normal=3) are downloaded from the GEO database. The ROC curves of the key characteristic genes in the two

diseases are drawn with the help of the “pROC” and “ggplot2” packages, and the AUC values and 95% CI values are calculated to verify the diagnostic efficacy of the key genes in the two diseases. Draw a violin plot using the t-test to verify the diagnostic value of key characteristic genes in the training sets GSE79768 and GSE37171.

2.7 Construction of gene-gene and TFs-gene interaction network diagrams

The gene-gene interaction network of biomarkers, that is, key characteristic genes, is analyzed with the help of the GeneMANIA online database; based on the JASPAR database (<https://jaspar.genereg.net/>), the transcription factors (TFs)-gene regulatory network is constructed using the NetworkAnalyst platform, and the key transcription factors are found by using the cytohubba plug-in MCC algorithm of Cytoscape software.

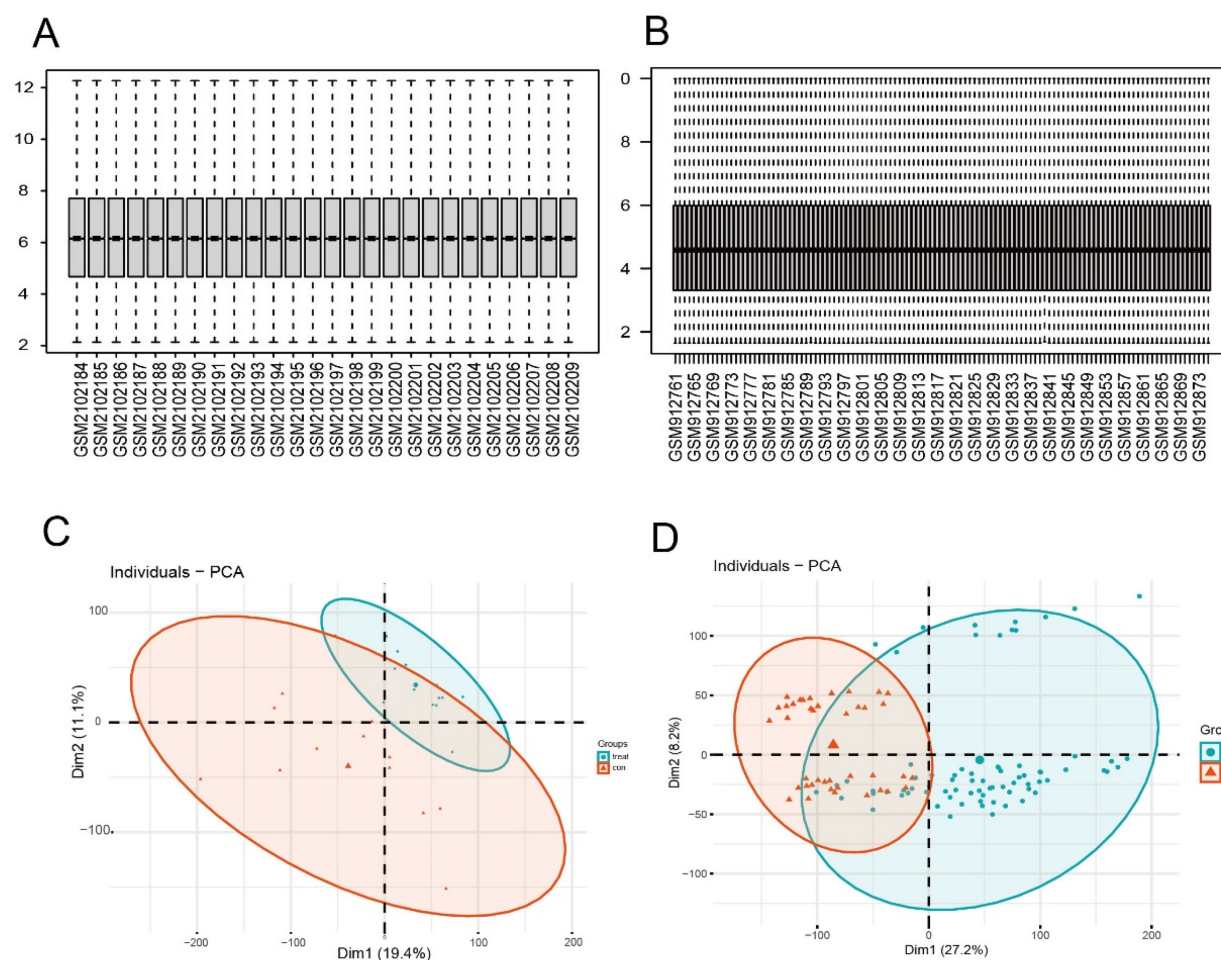


FIGURE 2
Normalized expression matrices. Box plots (A,B) and PCA diagrams (C,D) of the GSE79768 and GSE37171 datasets. PCA, principal component analysis.

2.8 Immune cell infiltration analysis

The expression data of 82 co-expressed differentially expressed genes in the atrial fibrillation validation set GSE79768 and the chronic kidney disease validation set GSE37171 were obtained. The infiltration abundances of immune cells in the two disease validation sets were evaluated based on the “CIBERSORT” R package. The results were displayed by drawing bar charts, box plots, and heat maps using ggplot2. The correlation coefficients and *P*-values between the target genes and immune cell subsets were calculated using the “cor.test” function, and the analysis results were visualized with lollipop plots.

3 Results

3.1 Screening and identification of DEGs of the two diseases

In the AF dataset GSE79768, which contains 14 atrial fibrillation patients and 12 sinus—rhythm normal control groups,

a total of 425 DEGs were screened (Figures 3A,B), among which 210 genes were up-regulated and 215 genes were down-regulated. In the CKD dataset GSE37171, which contains 75 chronic kidney disease patients and 40 normal control groups, a total of 4,128 DEGs were screened, among which 660 genes were up-regulated and 285 genes were down-regulated [with threshold conditions: *P*-value < 0.05 and $|\log_2(\text{FC})| > 1$, which were illustrated by the volcano plots in Figures 3C–E]. 82 co expressed DEGs were obtained by taking the intersection of the DEGs of the two diseases through Venn diagrams (Figure 4A).

3.2 GO/KEGG enrichment analyses of co-expressed DEGs

In the results of GO enrichment analysis, from the perspective of biological processes, the genes were mainly enriched in humoral immune responses, lymphocyte-mediated immunity, adaptive immune responses of immunological receptor somatic recombination based on the immunoglobulin superfamily domain, and myeloid leukocyte migration (Figures 4E,F). From

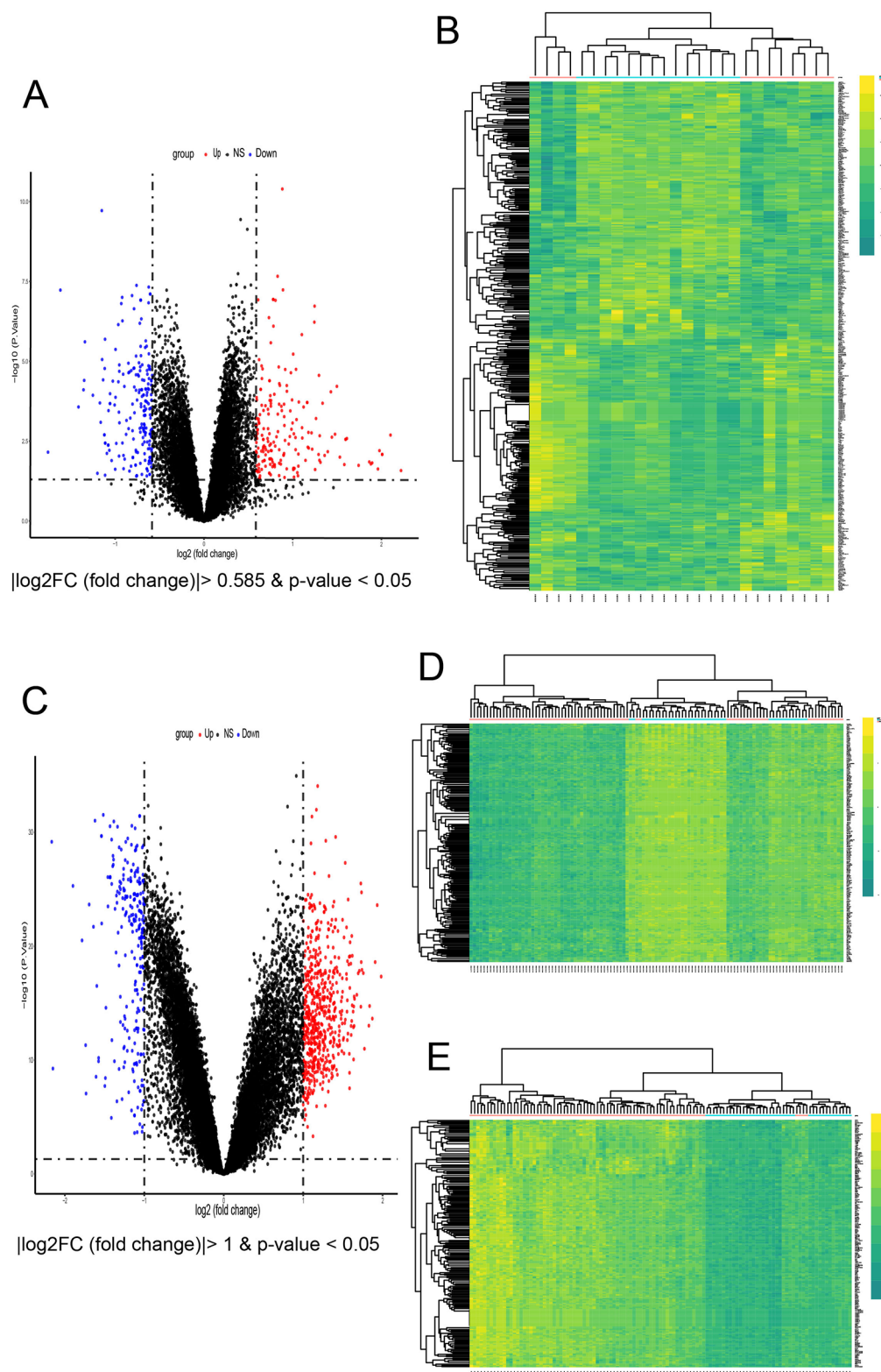


FIGURE 3

Identification of common DEGs and enrichment analysis. (A) Volcano plots of all DEGs in GSE79768. (B) Heatmap of all DEGs in GSE79768. (C) Volcano of all DEGs in GSE37171. (D,E) The top 200 up-regulated concurrently with the top 200 down-regulated DEGs of in GSE37171.

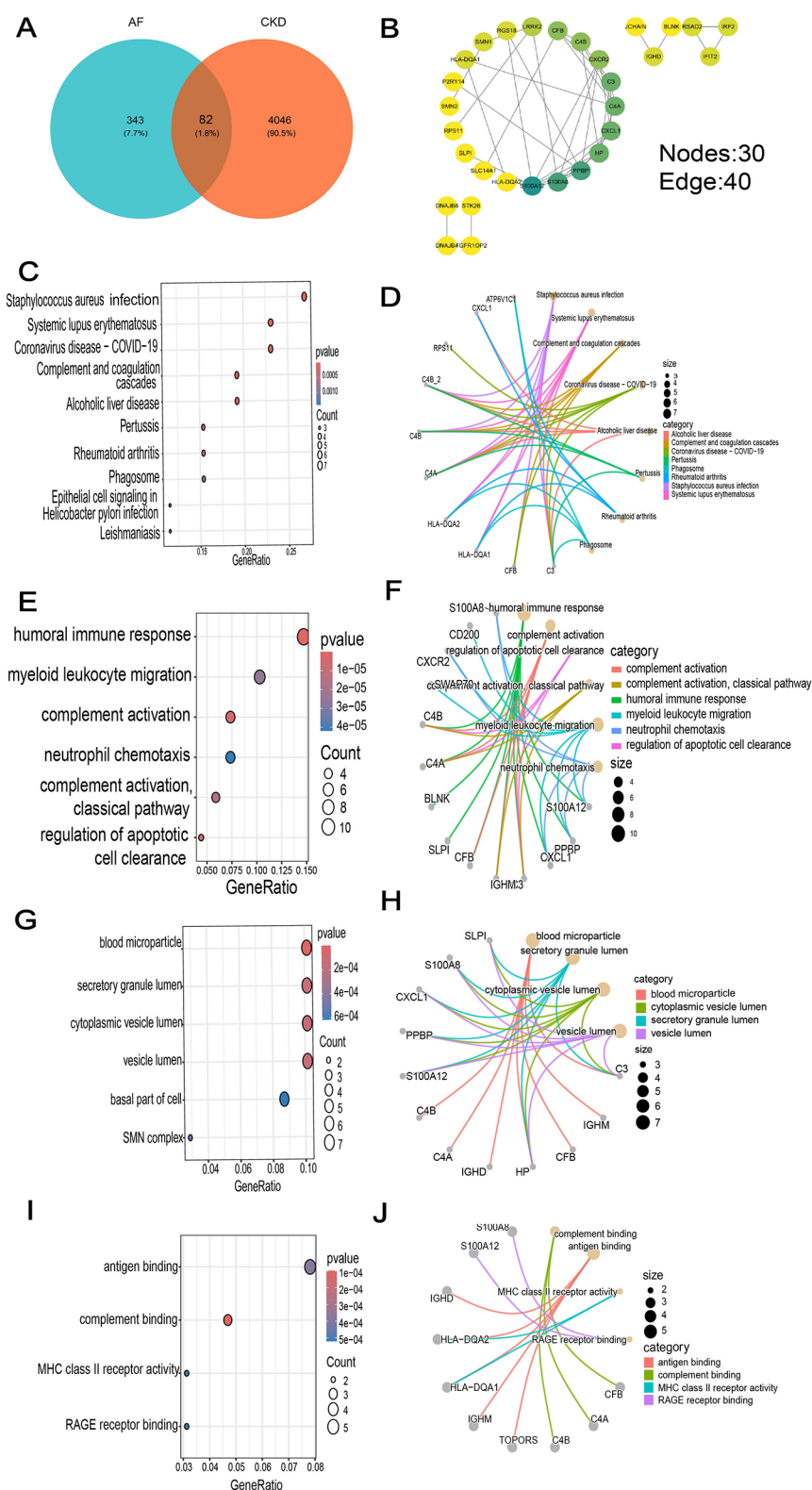


FIGURE 4

GO and KEGG enrichment analysis of common DEGs. (A) Venn diagram of common DEGs between AF and CKD. (B) The PPI network diagram beautified with the help of Cytoscape software. (C) Top 10 results in KEGG pathway enrichment analysis of the common DEGs. (D–J) The top 8 KEGG pathways (D) and the top 4 GO (BP, CC and MF) terms (E–J) of the common DEGs and their enriched targets were identified.

the perspective of cellular composition, the genes were mainly enriched in secretory granule lumen, blood microparticles, cytoplasmic vesicle lumen, and vesicle lumen (Figures 4G,H). From the perspective of molecular functions, the genes were mainly enriched in complement binding, antigen binding and MHC class II receptor activity (Figures 4I,J). KEGG functional enrichment analysis indicated that the genes were mainly enriched in *Staphylococcus aureus* infection, systemic lupus erythematosus, COVID-19, alcoholic liver disease, and the complement—coagulation cascade process (Figures 4C,D).

3.3 Constructing a PPI network to obtain candidate genes

A PPI network of 79 co-expressed genes was constructed using the STRING online website. After removing the isolated nodes, it was imported into Cytoscape software to obtain a PPI network diagram with 30 nodes and 40 edges (Figure 4B), and the top 15 genes were obtained based on the MCC algorithm (Figure 5A).

3.4 Feature genes are obtained by machine-learning algorithms

Further screening was performed on the top 15 genes of AF and CKD respectively based on two machine-learning algorithms (LASSO algorithm and RF algorithm). After gene screening for AF (Figures 5B,C) and CKD (Figures 5D,E) by the LASSO algorithm respectively, the intersection was taken to obtain 7 characteristic genes: PPBP, CXCL1, C3, LRRK2, RGS18, HLA—DQA1, RSAD2. When applying the RF algorithm for gene screening, with the criterion of Mean Decrease Gini >0.8 or 1 set, analyses were carried out for AF (Figure 5F) and CKD (Figure 5G) respectively, and the intersection of the two obtained through a Venn diagram yielded 7 characteristic genes: PPBP, CXCL1, CXCR2, LRRK2, RGS18, RSAD2, S100A8. Then, the characteristic genes obtained by the two machine-learning algorithms were intersected, and finally, 5 key characteristic genes were obtained: PPBP, CXCL1, LRRK2, RGS18, RSAD2 (Figure 5H).

3.5 Verification of gene diagnostic efficacy with external datasets

ROC curves were constructed for the five key characteristic genes in the external validation datasets respectively, and the diagnostic efficacy of the characteristic genes for diseases was verified by calculating the area under the curve. Verification in the AF dataset GSE14975 found that RSAD2 (AUC = 1.000, 95%CI = 1), CXCL1 (AUC = 0.880, 95%CI = 0.626–1), LRRK2 (AUC = 0.760, 95%CI = 0.4138–1), PPBP (AUC = 0.880, 95%CI = 0.626–1), and RGS18 (AUC = 0.720, 95%CI = 0.365–1), among which RSAD2 had the highest diagnostic value for AF (Figure 6A). Verification in the CKD dataset GSE120683 showed that RSAD2 (AUC = 0.889, 95%CI = 0.5809–1), CXCL1 (AUC = 0.889, 95%

CI = 0.5809–1), LRRK2 (AUC = 0.667, 95%CI = 0.1332–1), PPBP (AUC = 0.889, 95%CI = 0.5809–1), and RGS18 (AUC = 0.667, 95%CI = 0.1332–1), among which RSAD2, CXCL1, and PPBP had the same diagnostic value for CKD (Figure 6B). Therefore, three co-expressed biomarkers of the two diseases were obtained based on AUC > 0.7: PPBP, CXCL1, and RSAD2. In the two training sets of AF GSE79768 (Figure 6C) and CKD GSE37171 (Figure 6D), violin plots were drawn, and the results showed that the biomarkers had good diagnostic value ($p < 0.05$).

3.6 Construction of genes-genes and TFs-genes interaction network diagrams

The Genes-Genes interaction network of biomarkers was analyzed using the GeneMANIA database. The protein-interaction relationships involved in the three biomarkers are shown in Figure 7A, displaying the adjacent genes that change most frequently with the biomarkers. PPBP, CXCL1, and RSAD2 have complex interaction networks, and they interact with the surrounding shared genes through co-expression, physical interaction, prediction, genetic interaction, co-localization, pathways, and shared protein-domain interaction. Among them, the two biomarkers PPBP and CXCL1 mainly have functional interactions with cytokine activity, chemokine receptor activity, cell response to chemokines, neutrophil migration, response to chemokines, granulocyte chemotaxis, and granulocyte migration. The above results once again emphasize the importance of immune-cell-related inflammation in these two diseases, and PPBP and CXCL1 may be key regulatory genes in the development mechanisms of the two diseases.

The TFs-genes interaction network was constructed using the NetworkAnalyst platform to find potential regulatory molecules (Figure 7B), and the key transcription factors were found using the cytohubba plug-in MCC algorithm. Finally, it was confirmed that FOXC1, FOXL1, and GATA2 are the main transcription factors of the key characteristic genes (Figure 7C).

3.7 Correlation analysis between key characteristic genes and immune cell infiltration

Enrichment analysis of co-expressed genes showed that the genes were mainly enriched in pathways such as the complement-coagulation cascade and systemic lupus erythematosus, indicating that the genes are closely related to immune-inflammatory responses. To further explore the infiltration differences of various immune cells in the case and control groups of the two diseases respectively, the proportions of each sample of the two diseases in 22 kinds of immune cells were obtained based on the “CIBERSORT” algorithm (Figures 8A–D). The results showed that compared with healthy controls, the common feature of AF and CKD is a consistent upward trend of B memory cells and CD8+ T cells. In the AF dataset, compared with the control group, the AF case group was

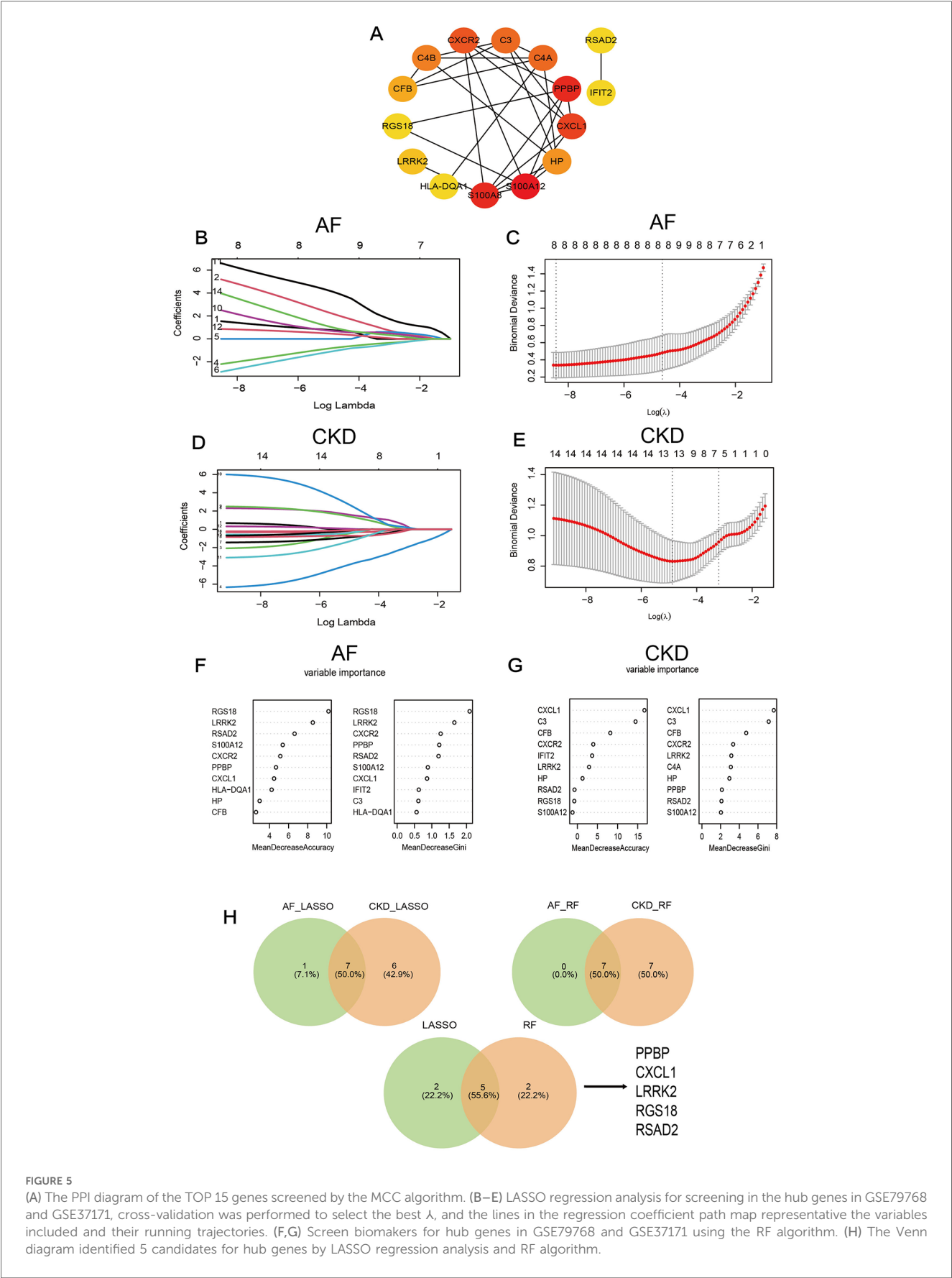


FIGURE 5 (A) The PPI diagram of the TOP 15 genes screened by the MCC algorithm. (B–E) LASSO regression analysis for screening in the hub genes in GSE79768 and GSE37171, cross-validation was performed to select the best λ , and the lines in the regression coefficient path map representative the variables included and their running trajectories. (F,G) Screen biomarkers for hub genes in GSE79768 and GSE37171 using the RF algorithm. (H) The Venn diagram identified 5 candidates for hub genes by LASSO regression analysis and RF algorithm.

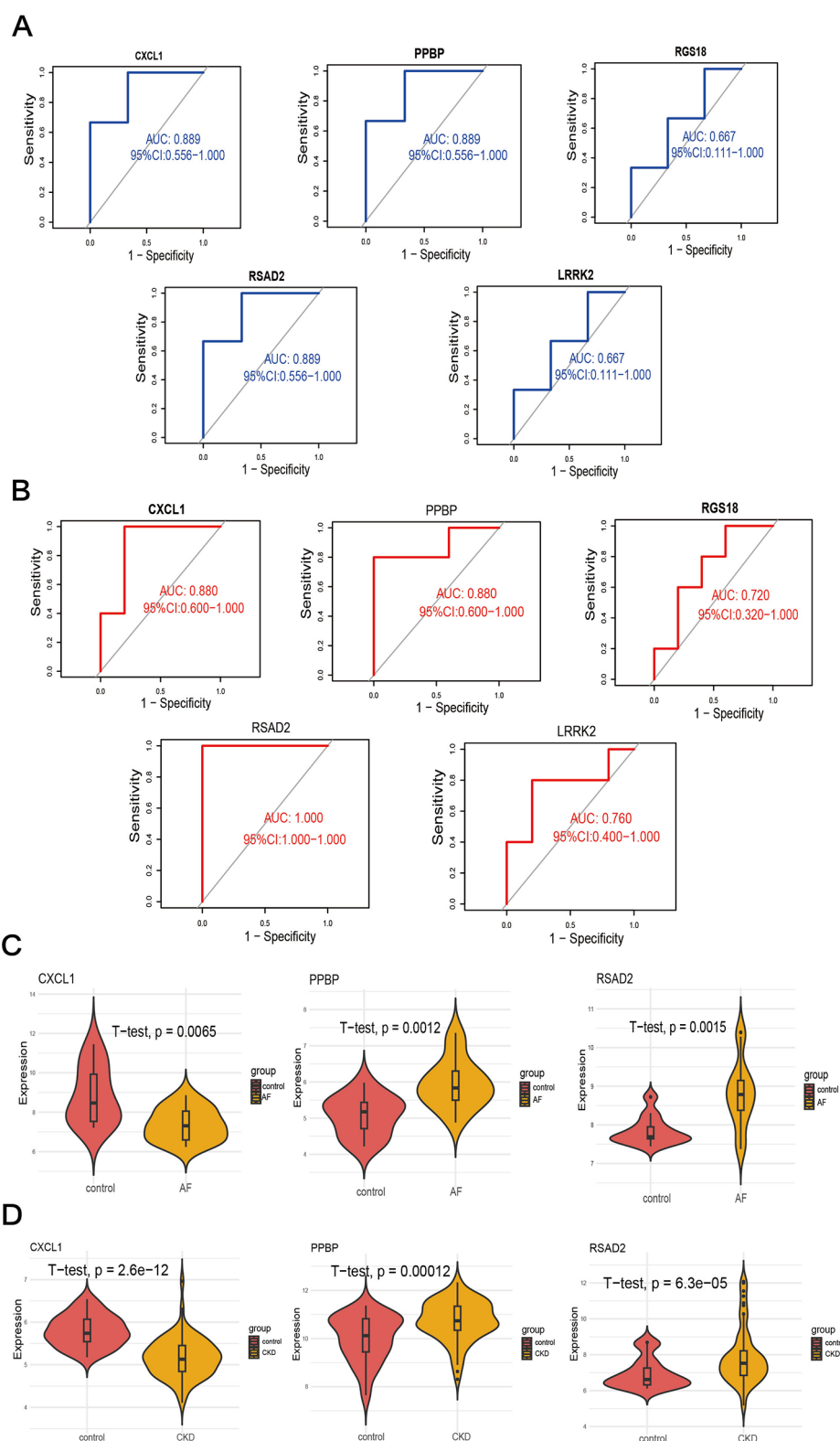


FIGURE 6

Validation of candidate genes. (A,B) ROC curves of candidates genes to assess for diagnostic sensitivity and specificity value in GSE14975 (A) and GSE120683 (B). (C,D) Validation of the diagnostic efficacy of candidate genes in the GSE79768 (C) and GSE37171 (D) respectively by *T*-test.

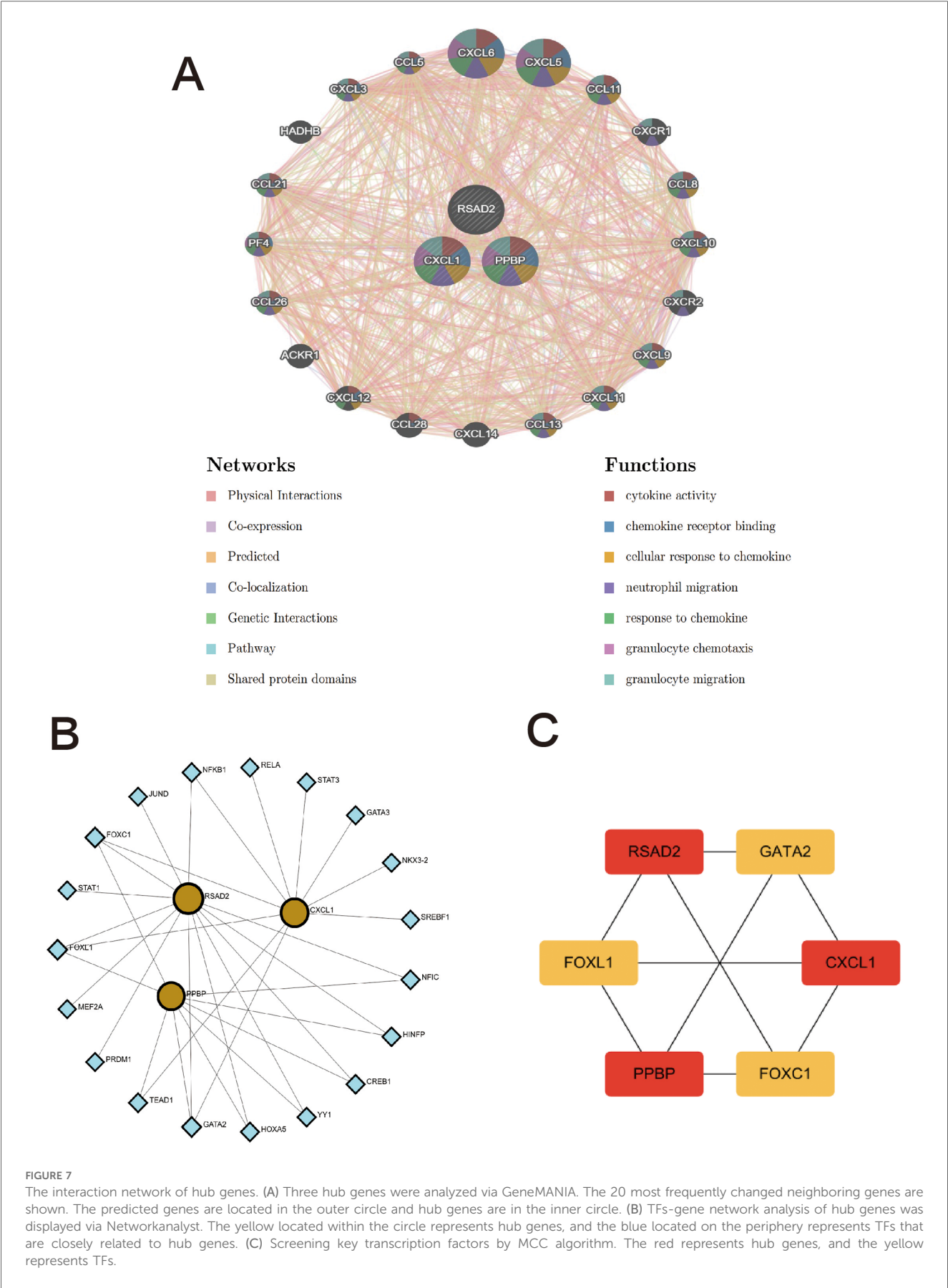


FIGURE 7
The interaction network of hub genes. (A) Three hub genes were analyzed via GeneMANIA. The 20 most frequently changed neighboring genes are shown. The predicted genes are located in the outer circle and hub genes are in the inner circle. (B) TFs-gene network analysis of hub genes was displayed via NetworkAnalyst. The yellow located within the circle represents hub genes, and the blue located on the periphery represents TFs that are closely related to hub genes. (C) Screening key transcription factors by MCC algorithm. The red represents hub genes, and the yellow represents TFs.

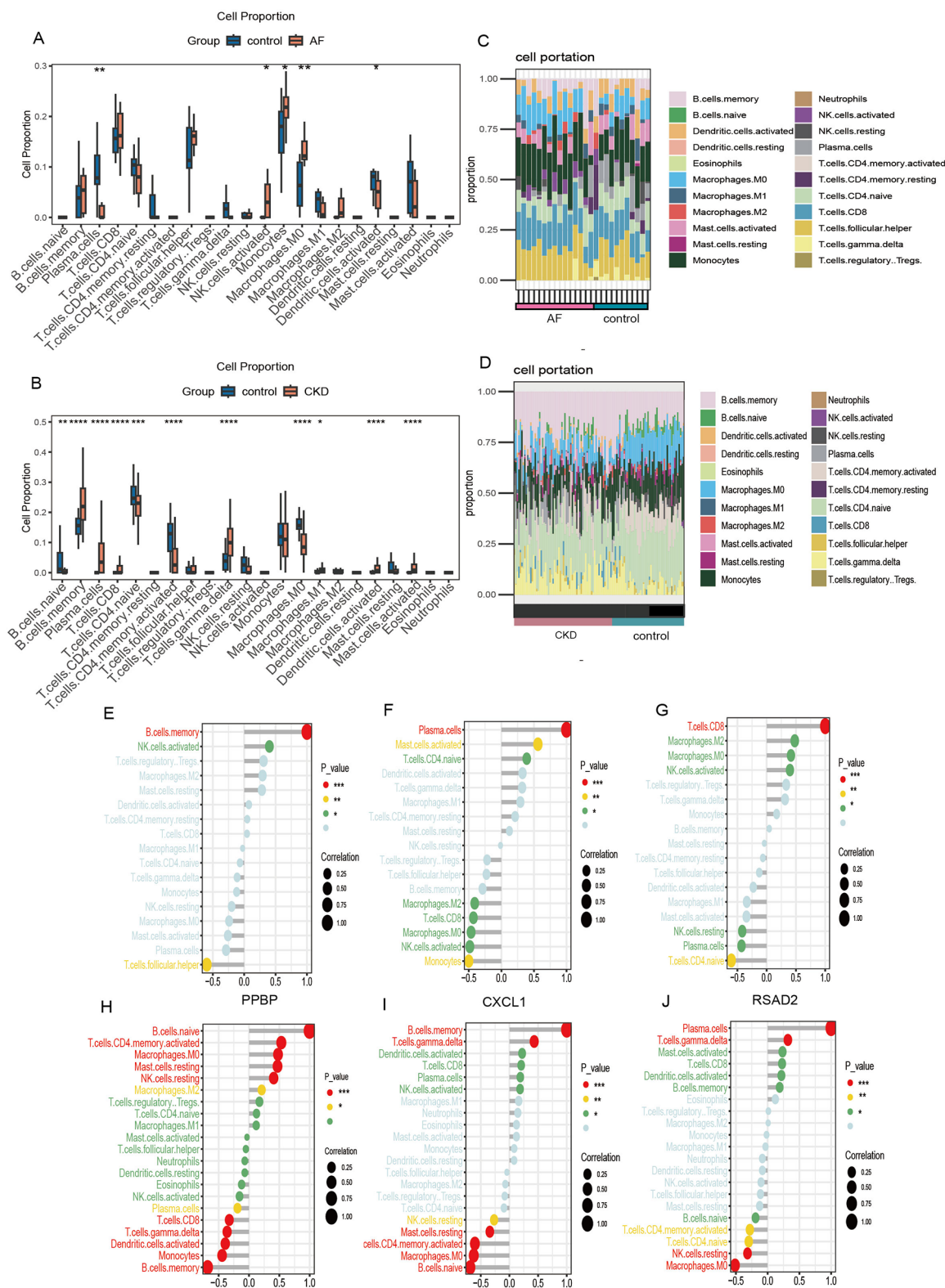


FIGURE 8 Immune cell infiltration. (A,B) Box plots show the comparison of 22 types of immune cells between the disease group and the control group of GSE79768 and GSE37171. (C,D) Immune cell infiltration maps of each sample of GSE79768 and GSE37171. (E–G) The correlations between the expression of three hub genes (PPBP, CXCL1, RSAD2) and immune cell in GSE79768. (H–J) The correlations between the expression of three hub genes (PPBP, CXCL1, RSAD2) and immune cell in GSE37171. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

significantly up-regulated in monocytes, M0 macrophages and activated natural killer (NK) cells ($p < 0.05$), and significantly down-regulated in activated dendritic cells and plasma cells ($p < 0.05$). In the CKD dataset, the CKD case group was significantly up-regulated in B memory cells, γ - δ T cells, plasma cells, activated mast cells, M1 macrophages, activated dendritic cells, and CD8+ T cells ($p < 0.05$), and significantly down-regulated in resting CD4+ T cells, M0 macrophages, activated CD4+ memory cells, and resting B cells ($p < 0.05$).

To further investigate the association between biomarkers and immune cells, Spearman rank analysis was carried out. In the figure, the horizontal axis represents the correlation between biomarkers and immune cells, with dots representing the size of the correlation coefficient. The larger the dot, the greater the correlation. The vertical axis represents the types of immune cells, and the number of asterisks indicates the size of the P -value. The smaller the P -value, the more asterisks.

The results of the correlation analysis show that in the AF data, PPBP is significantly positively correlated with B memory cells and activated NK cells, and significantly negatively correlated with follicular helper T cells (Figure 8E); CXCL1 is significantly positively correlated with plasma cells, activated mast cells, and resting CD4+ T cells, and significantly negatively correlated with monocytes, activated NK cells, CD8+ T cells, M0- and M2- type macrophages (Figure 8F); RSAD2 is significantly positively correlated with CD8+ T cells, activated NK cells, M0- and M2- type macrophages, and significantly negatively correlated with resting CD4+ T cells, plasma cells, and resting NK cells (Figure 8G). In the CKD data, PPBP is significantly positively correlated with resting B cells, activated CD4+ T memory cells, resting mast cells, resting NK cells, M0- and M2- type macrophages, and significantly negatively correlated with B memory cells, monocytes, activated dendritic cells, γ - δ T cells, CD8+ T cells, and plasma cells (Figure 8H); CXCL1 is significantly positively correlated with memory B cells and γ - δ T cells, and significantly negatively correlated with resting B cells, M0- type macrophages, activated CD4+ T memory cells, resting mast cells, and resting NK cells (Figure 8I); RSAD2 is significantly positively correlated with plasma cells and γ - δ T cells, and significantly negatively correlated with M0- type macrophages, resting NK cells, activated CD4+ T memory cells, and resting CD4+ T cells (Figure 8J).

4 Discussion

At present, an increasing number of studies have shown that the systemic inflammatory response state and immune cells play an important role in the occurrence and development of AF and CKD. For example, mast cells promote the expression of cardiac fibroblast cells and collagen by secreting platelet-derived growth factor-A (PDGF A) (14). During heart injury, macrophages present antigens released by dead heart tissue cells to CD4+ T cells through protein molecules on their surface and activate them, leading to the continuous release of inflammatory factors and the continuous progress of pro-fibrosis, thereby expanding

heart damage (15–17). When the kidneys are damaged, it causes the continuous release of inflammatory factors, resulting in continuous damage to kidney tissue and function, thereby leading to CKD. Some studies have pointed out after ureteral ligation in wild-type mice, disodium cromoglycate, a mast cell stabilizer that reduces degranulation, significantly alleviated renal fibrosis. This conclusion is consistent with the immune infiltration results of CKD disease in this study (18). Similarly, both AF and CKD are two diseases with a high risk of thrombosis. Both are associated with an increased risk of thrombosis and embolism events. The process of thrombosis involves the participation of various cytokines and coagulation factors, which play important roles. A study exploring a mouse model found that edoxaban significantly inhibited the levels of inflammatory factors, upregulated the levels of chemokines, and reduced the risks of apoptosis, atrial fibrillation, and venous thrombosis by regulating the Wnt- β -mediated PI3K/AKT signaling pathway (19). In this study, through KEGG enrichment analysis, it was found that the differentially expressed genes were enriched in the complement—coagulation cascade process, which is consistent with the research findings. In addition, we found that CXCL1 and PPBP are closely associated with CXCR2 (C-X-C motif chemokine receptor 2). As a chemokine receptor, CXCR2 can mediate apoptosis, inflammatory responses, and angiogenesis through pathways such as the phosphatidylinositol 3-kinase/Akt (PI3K/Akt) pathway, the phospholipase C/protein kinase C (PLC/PKC) pathway, and the NF- κ B pathway. These processes are closely related to the mechanism of thrombosis.

In our study, we attempted to identify DEGs and deeply explore the role of immune cell infiltration in AF-CKD. A total of 82 DEIRGs were detected as candidate biomarkers in the pathogenesis of AF-CKD, and the enrichment function identified the potential mechanisms of DEIRGs. The results of KEGG enrichment analysis show that DEIRGs are significantly related to immune and inflammatory responses such as systemic lupus erythematosus, COVID-19, and the complement- coagulation cascade. In addition, through GO analysis, these genes are involved in humoral immune responses, lymphocyte- mediated immunity, adaptive immune responses of immunological receptor somatic recombination based on the immunoglobulin superfamily domain, secretory granule lumen, blood microparticles, cytoplasmic vesicle lumen, vesicle lumen, complement binding, antigen binding, and other pathways. The study found three potential biomarkers of AF and CKD-PPBP, CXCL1, and RSAD2, and the CXCLs/CXCR signaling pathway plays an important role in the occurrence and development of the two diseases.

CXCL1 (C-X-C motif chemokine receptor 1) is a member of the CXC superfamily and an important chemokine that participates in the occurrence and development of various inflammatory diseases, activating CXC chemokine receptor 2 (CXCR2) and high-level CXC chemokine receptor 1 (CXCR1). It has important functional roles in physiological functions, including inducing angiogenesis and neutrophil recruitment. In a 2018 study by Lei Wang et al. (20), the researchers found that angiotensin II-induced monocyte infiltration in the heart is mainly mediated by CXCL1-CXCR2 signaling, and

CXCL1-CXCR2 signaling initiates and promotes cardiac remodeling. In addition, the study pointed out that intervention with CXCL1 neutralizing antibody prevents the development of angiotensin II-induced cardiac hypertrophy, fibrosis, and inflammation (21). As a pro-inflammatory chemokine, CXCL1 mediates various acute and chronic inflammations, thereby promoting fibrosis development. In particular, monocyte infiltration in atrial tissue leads to an increase in the number of macrophages in the atrium, and then macrophages can directly couple with surrounding cardiomyocytes through gap junctions and increase atrioventricular conduction by stimulating cardiomyocyte repolarization (22). The expression level of CXCL1 in the atrium of atrial fibrillation patients is strongly positively correlated with activated mast cells (23), which is consistent with the results of immune infiltration analysis in this study. Activated mast cells and macrophages (24), especially those interacting with vascular smooth muscle cells, produce CXCL1 in the intima and perivascular tissue. In AF patients (14), mast cells may induce cardiac fibroblasts to increase collagen production by secreting PDGF-A, resulting in atrial fibrosis and AF. CXCL1 may mediate inflammatory responses through the NF- κ B (nuclear factor- κ B), TGF-Smad 2/3 signaling pathways, among which TGF- β is an important pro-fibrotic cytokine. Similarly, some studies have pointed out (25) that in damaged kidney cells, when TGF- β is continuously expressed, pathological accumulation of extracellular matrix substances in the glomerulus and interstitial compartments occurs, leading to the progression of renal fibrosis (26). Damaged renal tubules activate myofibroblast differentiation, proliferation, and matrix secretion by producing and releasing bioactive molecules to recruit inflammatory cells. A single-cell analysis pointed out (27) that CXCL1 is secreted from pro-fibrotic tubules and then recruits CXCR2+ basophils, which is a key factor in renal fibrosis. Some studies have found that the expression of TNF- α , and CXCL-1 increases in the kidneys of CKD rats. After treatment with tanshinone IIA, it was found that the drug inhibits kidney inflammation mediated by the NF- κ B signal and reduces the expression of these inflammatory genes (26).

PPBP (pro-platelet basic protein), also known as CXCL7 or NAP2, belongs to the platelet-derived growth factors of the CXC chemokine family. It is an effective chemoattractant and activator of neutrophils. It is released from activated platelets and participates in the response to vascular injury. It can stimulate various processes, such as mitosis, glucose metabolism, and the synthesis of extracellular matrix and plasminogen activators. Some studies have pointed out (23) that CXCL7/PPBP and CXCL1 are the hub genes with the most associations in atrial fibrillation. It is now clear that PPBP plays a crucial role in coordinating neutrophil recruitment during vascular injury. Some studies have explored (28, 29) the molecular basis for chemokine CXCL7 to form heterodimers with chemokines CXCL1, CXCL4, and CXCL8, pointing out that these chemokines co-exist in platelet granules and are released upon platelet activation. By activating the CXCR2 receptor, they recruit and coordinate neutrophils to reach the site of tissue injury. The recruited neutrophils mediate microbial killing and initiate tissue repair.

PPBP exerts the functions of recruiting neutrophils and inducing angiogenesis by activating the CXCR2 receptor and binding to sulfated glycosaminoglycans (sGAGs). However, continuously activated neutrophils further aggravate the initial injury, leading to various thrombosis-related cardiovascular and inflammatory diseases (30–33). An article published in 2013 pointed out (34) that there is a CXCL7 chemotactic gradient within the thrombus, which is indispensable for the migration of neutrophils within the thrombus. In addition, the study (35) showed that in the context of platelet-neutrophil communication, the ability of CXCL7 to exist as both a monomer and a dimer, as well as in both free and GAG-bound states, may maximize repair and reduce tissue injury and the occurrence of cardiovascular diseases. CXCL7 is unique.

RSAD2 (radical S-adenosyl methionine domain containing 2), also known as Viral inhibitory protein, Endoplasmic reticulum-associated, interferon-inducible (VIPERIN), is an antiviral gene. In a recent study by Örebro University (36), publicly available human carotid atherosclerosis data were used and experimental research was carried out. The results showed that the RSAD2 gene has a higher expression level in plaques. Further experimental research found that when human-derived aortic smooth muscle cells are stimulated by interferon- γ , knocking down the RSAD2 gene will lead to a decrease in the expression and secretion of CXCR3 chemokines CXCL9, CXCL10, and CXCL11, thereby reducing the chemotactic effect on monocytes and decreasing the migration of monocytes to the lesion site. During the process of atherosclerosis, RSAD2 may play a role through inflammatory and immune responses. In addition, studies have shown (37) that functional chemokine receptors, including CXCR3, are expressed in glomerular podocytes, and the activation of these receptors may induce the release of oxygen free radicals, leading to podocyte injury and the development of proteinuria. Currently, there are relatively few studies on RSAD2, and more research is needed to further explore the potential role of this gene in various diseases.

In the studies of the three biomarkers, it is summarized that PPBP, CXCL1, and RSAD2 have similarities in the disease development process. All of them can lead to the up-regulated CXCLs inducing, activating, and binding to CXCR chemokines, promoting the rapid recruitment of leukocytes in the inflammatory environment. Therefore, the CXCLs/CXCR signaling pathway may play an important role in the development of AF and CKD. The research on the CXCLs/CXCR signaling pathway in inflammatory diseases can involve many different systems, including the respiratory system, digestive system, cardiovascular system, and nervous system. The development of diseases in different systems is inseparable from extensive leukocyte infiltration. After receiving upstream signaling, chemokine receptors recruit leukocytes to gather at the inflammatory foci. However, leukocyte disorders and excessive infiltration lead to the destruction of organ structures, resulting in inflammation-related diseases. Among them, CXCR2, a chemokine receptor closely related to PPBP and CXCL1 in the development of AF and CKD, is a G-protein-coupled receptor (GPCR) and plays an important role in the human immune

system, especially in cell chemotaxis, inflammatory responses, and tumorigenesis. Binding to CXCR2 can activate multiple G-protein—mediated signal cascade pathways, including the PI3K/Akt pathway, PLC/PKC pathway, mitogen-activated protein kinase/p38 (MAPK/p38) pathway, and Janus kinase 2/signal transducer and activator of transcription (JAK2/STAT3) pathway. These signal transduction pathways are several key intracellular signal transduction routes, and they play important roles in cell metabolism, apoptosis, inflammatory responses, angiogenesis, cytokine signaling, and immune responses (38). Different from PPBP and CXCL1, the expression of RSAD2 is related to CXCR3 chemokines. CXCR3 is one of the G-protein-coupled receptors, mainly binding to ligands such as CXCL9, CXCL10, and CXCL11, and participating in the chemotaxis of immune cells.

In addition, this study constructed a TFs-DEGs interaction network for further analysis of possible therapeutic targets. In the TFs-DEGs interaction network, FOXC1, FOXL1, and GATA2 were identified as key transcription factors. FOXC1 and FOXL1 transcription factors both belong to the “C” subgroup of the Forkhead box (FOX)/Winged helix family. FOX transcription factors play important roles in various biological processes, including cell proliferation, differentiation, survival, and death. FOX transcription factors are distributed in different tissues and cell types of the lungs, intestines, and cardiovascular system (39). They are involved in cell proliferation, autophagy, inflammation, and the differentiation of specific gene transcription during the development and adulthood of cardiovascular tissues (40, 41). In a human heart tissue experimental study (42), it was found that compared with the control group of human-induced pluripotent stem-cell-derived cardiomyocytes with normal telomere length (nTL-CMs), human-induced pluripotent stem-cell-derived cardiomyocytes with short telomeres (sTL-CMs) exhibited an increase in aging markers, a decrease in mitochondrial respiratory reserve, and contractile dysfunction. In addition, the study found that overexpression of FOXC1 in nTL-CMs fully induces myocardial aging, induces inflammation—related genes, and simultaneously down-regulates myofibrillar and mitochondrial genes, leading to contractile dysfunction and myocardial aging. Conversely, knockdown of shFOXC1 in sTL-CMs is sufficient to block aging and restore mitochondrial function. Wataru et al. found that in patients with AF, the long non-coding RNA (lncRNA) FOXCUT strongly upregulates the expression of the transcription factor FOXC1. It exerts a post-transcriptional genetic regulation similar to that in carcinogenesis, prompting the transformation of endothelial cells into fibroblasts. This leads to atrial structural remodeling and the occurrence of arrhythmia (43). It has been pointed out (44) that FOXL1 can regulate a series of genes that enhance fibroblast function, including the transcriptional co-activator Yes-associated protein and/PDZ-binding motif (YAP/TAZ) marker genes and platelet-derived growth factor receptor- α . YAP/TAZ are downstream effectors of the Hippo signaling pathway and are crucial transcriptional co-activators. Some research reports that the FOX family is expressed in renal tubular epithelial cells during renal fibrosis and exhibits the ability to promote fibrosis and metabolic rewiring (45). The GATA2 transcription factor is

a member of the GATA-binding transcription factor family and plays a key role in biological processes such as hematopoiesis and endocrine cell development, cell proliferation, and immune responses. For example, in terms of immune responses, GATA2 is involved in regulating the differentiation of dendritic cells, the proliferation of hematopoietic stem cells, the differentiation of IgE/mast cells, and the differentiation of granulocyte—monocyte progenitor cells (46, 47). Some studies have found that FOG2 is overexpressed in human heart failure. In addition, conditional overexpression of FOG2 in adult mouse cardiomyocytes leads to the occurrence of primary AF. As a protein that can bind directly to GATA2, FOG2 can cooperate with GATA2 to regulate the expression of target genes after they bind to each other (48). Studies have shown (49) that overexpression of GATA2 reduces the phagocytic ability of phagocytes and affects the maturation of phagosomes. A mouse model study found (50) that the expression of GATA2 in mice can enhance the production of inflammatory mediators by adipocytes, and in mouse macrophages, the toll-like receptor-4 (TLR4) signaling pathway up-regulates GATA2, and subsequently GATA2 cooperates with the mitogen-activated protein kinase (MAPK) signaling pathway to induce IL-1 β transcription. Interestingly, YAP has been found to interact with the transcription factor GATA2, enhancing the transcription of aquaporin-2 (AQP2) in the renal collecting ducts, and controlling the final urine volume and osmotic pressure. When the YAP gene is deficient, the mRNA and protein abundance of Aqp2 in the kidneys is significantly reduced, leading to disorders in renal water—salt metabolism and further promoting the progression of kidney diseases (51). Some scholars have proposed that in kidney injury caused by ischemia-reperfusion injury, GATA2 positively regulates the expression of cytokines and chemokines in collecting duct cells, promoting disease progression. By inhibiting GATA2, the expression of inflammatory factors can be reduced, and kidney damage can be significantly alleviated (52).

Our study has discovered that PPBP, CXCL1, and RSAD2 promote the development of AF and CKD, among which the CXCLs/CXCR signaling pathway may play a crucial role. In addition, the related transcription factors FOXC1, FOXL1, and GATA2 also lead to disease progression by mediating the high expression of inflammatory factors, profibrotic factors, and immune cells. Based on the current research findings on the potential pathogenesis of these two diseases, these discoveries provide guiding suggestions for us to develop more effective treatment plans in clinical practice. Currently, in the standardized treatment of AF, the focus is mainly on three aspects: rhythm control, heart rate control, and anticoagulation therapy. Among them, pulmonary vein isolation (PVI) is a widely recognized treatment for rhythm control and serves as the cornerstone of catheter ablation for AF. Vinitha Nesapiragasam et al. (53) pointed out that even with persistent PVI, there is still a possibility of arrhythmia recurrence. For patients with ineffective PVI or recurrence after persistent PVI, catheter ablation strategies other than PVI can be adopted. Among these, renal denervation (RDN) can reduce the frequency of AF. In a study on a canine model by Wang et al. (54), the results showed

that RDN inhibited the upward trend of TNF- α and IL-6 levels in the left atrium after prolonged right atrial pacing, and significantly alleviated the damage of myocardial fibrosis caused by chronic right atrial pacing. This demonstrated that catheter-based RDN could suppress the activity of the RAAS and atrial remodeling, including the significant increase in atrial fibrosis, inflammation, and apoptosis. In a randomized controlled study, it was found that among patients with paroxysmal AF and hypertension, compared with using catheter ablation alone, adding RDN significantly reduced the likelihood of AF recurrence (55). In addition, some studies have pointed out (56) that RDN inhibits the over-activation of the systemic sympathetic nervous system, providing strong protection against the occurrence and progression of kidney and heart diseases. For patients with both AF and CKD, in addition to conventional drug therapies, opting for RDN might be a favorable treatment option. But at present, RND mainly focuses on the treatment of hypertension (57). There are relatively few studies on using RND to treat AF and CKD. Larger-scale and more in-depth research needs to be carried out. However, the broad prospects of the clinical application of RND can be anticipated.

In summary, in the pathogenesis of AF and CKD, chronic inflammation can increase the risk of AF and CKD through promoting atrial remodeling, oxidative stress, fibrosis and other pathways. PPBP, CXCL1 and RSAD2 may play important roles in the occurrence and development of AF and CKD, among which the CXCLs/CXCR signaling pathway may play a crucial role. With a better understanding of the underlying mechanisms of AF and CKD, patients with AF-CKD will be better managed and treated.

4.1 Limitation

This study has certain limitations. First, due to the diverse sources, incompleteness and multi-sourced nature of biological data, during the data processing, the identification of biomarkers may be interfered by batch effects and missing data values. Second, in the analysis of gene sequencing data, phenomena such as gene horizontal transfer and mutation can lead to inaccurate results and over-fitting of the model, affecting the reliability of the model. Finally, due to the complexity of biological systems, the limitations of data analysis methods and the uncertainty of biological significance, false—positive or false—negative results may occur. Therefore, the specific functions and mechanisms may need to be further verified through experiments.

5 Conclusion

In this study, three biomarkers (PPBP, CXCL1, and RSAD2) were identified as diagnostic markers for AF-CKD patients. It was found that immune and inflammatory pathways accelerate the occurrence and development of AF-CKD, among which the CXCLs/CXCR signaling pathway may play a crucial role. FOXCL1, FOXL1, and GATA2 may be potential therapeutic targets for AF

combined with CKD. The correlations between PPBP, CXCL1, RSAD2, and immune cells may play important roles in the pathogenesis of AF-CKD. Whether these biomarkers can become new targets for molecular—targeted therapy in AF-CKD patients requires further research.

Data availability statement

The original contributions presented in the study are publicly available. These data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc>, accession numbers: GSE79768, GSE37171, GSE120683, GSE14975.

Author contributions

TJ: Data curation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. GD: Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by Guangxi Key Clinical Specialty Construction Project, the Scientific Research Project of the Guangxi Health Commission (Z20201226) and the funding from the National Natural Science Foundation of China (No. 82260093).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Kornej J, Börschel CS, Benjamin EJ, Schnabel RB. Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights. *Circ Res.* (2020) 127(1):4–20. doi: 10.1161/CIRCRESAHA.120.316340
- Deng H, Guo P, Zheng M, Huang J, Xue Y, Zhan X, et al. Epidemiological characteristics of atrial fibrillation in southern China: results from the Guangzhou heart study. *Sci Rep.* (2018) 8(1):17829. doi: 10.1038/s41598-018-35928-w
- Wang Z, Chen Z, Wang X, Zhang L, Li S, Tian Y, et al. The disease burden of atrial fibrillation in China from a national cross-sectional survey. *Am J Cardiol.* (2018) 122(5): 793–8. doi: 10.1016/j.amjcard.2018.05.015
- Kalenderoglu K, Hayiroglu MI, Cinar T, Oz M, Bayraktar GA, Cam R, et al. Comparison of inflammatory markers for the prediction of atrial fibrillation recurrence following cryoablation. *Biomark Med.* (2024) 18(17–18): 717–25. doi: 10.1080/17520363.2024.2395236
- Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet.* (2017) 389(10075):1238–52. doi: 10.1016/s0140-6736(16)32064-5
- Lameire NH, Levin A, Kellum JA, Cheung M, Jadoul M, Winkelmayer WC, et al. Harmonizing acute and chronic kidney disease definition and classification: report of a kidney disease: improving global outcomes (KDIGO) consensus conference. *Kidney Int.* (2021) 100(3):516–26. doi: 10.1016/j.kint.2021.06.028
- Glasscock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls. *Nat Rev Nephrol.* (2017) 13(2):104–14. doi: 10.1038/nrneph.2016.163
- Graterol Torres F, Molina M, Soler-Majoral J, Romero-González G, Rodríguez Chitiva N, Troya-Saborido M, et al. Evolving concepts on inflammatory biomarkers and malnutrition in chronic kidney disease. *Nutrients.* (2022) 14(20):4297. doi: 10.3390/nu14204297
- Fu Y, Xiang Y, Li H, Chen A, Dong Z. Inflammation in kidney repair: mechanism and therapeutic potential. *Pharmacol Ther.* (2022) 237:108240. doi: 10.1016/j.pharmthera.2022.108240
- Bock JS, Gottlieb SS. Cardiorenal syndrome: new perspectives. *Circulation.* (2010) 121(23):2592–600. doi: 10.1161/CIRCULATIONAHA.109.886473
- Ronco C, House AA, Haapio M. Cardiorenal syndrome: refining the definition of a complex symbiosis gone wrong. *Intensive Care Med.* (2008) 34(5):957–62. doi: 10.1007/s00134-008-1017-8
- Rangaswami J, Bhalla V, Blair JEA, Chang TI, Costa S, Lentine KL, et al. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the American Heart Association. *Circulation.* (2019) 139(16):e840–78. doi: 10.1161/CIR.0000000000000664
- Siragy HM, Carey RM. Role of the intrarenal renin-angiotensin-aldosterone system in chronic kidney disease. *Am J Nephrol.* (2010) 31(6):541–50. doi: 10.1159/000313363
- Liao CH, Akazawa H, Tamagawa M, Ito K, Yasuda N, Kudo Y, et al. Cardiac mast cells cause atrial fibrillation through PDGF-A-mediated fibrosis in pressure-overloaded mouse hearts. *J Clin Invest.* (2010) 120(1):242–53. doi: 10.1172/JCI39942
- Lafuse WP, Wozniak DJ, Rajaram MVS. Role of cardiac macrophages on cardiac inflammation, fibrosis and tissue repair. *Cells.* (2021) 10(1):51. doi: 10.3390/cells10010051
- Wagner MJ, Khan M, Mohsin S. Healing the broken heart; the immunomodulatory effects of stem cell therapy. *Front Immunol.* (2020) 11:639. doi: 10.3389/fimmu.2020.00639
- Hofmann U, Beyersdorf N, Weirather J, Podolskaya A, Bauersachs J, Ertl G, et al. Activation of CD4+ T lymphocytes improves wound healing and survival after experimental myocardial infarction in mice. *Circulation.* (2012) 125(13):1652–63. doi: 10.1161/CIRCULATIONAHA.111.044164
- Summers SA, Gan PY, Dewage L, Ma FT, Ooi JD, O'Sullivan KM, et al. Mast cell activation and degranulation promotes renal fibrosis in experimental unilateral ureteric obstruction. *Kidney Int.* (2012) 82(6):676–85. doi: 10.1038/ki.2012.211
- Shan X, Liu Z, Wulasihan M, Ma S. Edoxaban improves atrial fibrillation and thromboembolism through regulation of the wnt- β -induced PI3K/ATK-activated protein C system. *Exp Ther Med.* (2019) 17(5):3509–17. doi: 10.3892/etm.2019.7379
- Wang L, Zhang YL, Lin QY, Liu Y, Guan XM, Ma XL, et al. CXCL1-CXCR2 axis mediates angiotensin II-induced cardiac hypertrophy and remodeling through regulation of monocyte infiltration. *Eur Heart J.* (2018) 39(20):1818–31. doi: 10.1093/eurheartj/ehy085
- Wu CL, Yin R, Wang SN, Ying R. A review of CXCL1 in cardiac fibrosis. *Front Cardiovasc Med.* (2021) 8:674498. doi: 10.3389/fcvm.2021.674498
- Hulsmans M, Clauss S, Xiao L, Aguirre AD, King KR, Hanley A, et al. Macrophages facilitate electrical conduction in the heart. *Cell.* (2017) 169(3):510–22.e20. doi: 10.1016/j.cell.2017.03.050
- Li S, Jiang Z, Chao X, Jiang C, Zhong G. Identification of key immune-related genes and immune infiltration in atrial fibrillation with valvular heart disease based on bioinformatics analysis. *J Thorac Dis.* (2021) 13(3):1785–98. doi: 10.21037/jtd-21-168
- Tucureanu MM, Butoi E, Gan AM, Stan D, Constantinescu CA, Calin M, et al. Amendment of the cytokine profile in macrophages subsequent to their interaction with smooth muscle cells: differential modulation by fractalkine and resistin. *Cytokine.* (2016) 83:250–61. doi: 10.1016/j.cyto.2016.04.019
- Gao L, Zhong X, Jin J, Li J, Meng XM. Potential targeted therapy and diagnosis based on novel insight into growth factors, receptors, and downstream effectors in acute kidney injury and acute kidney injury-chronic kidney disease progression. *Signal Transduct Target Ther.* (2020) 5(1):9. doi: 10.1038/s41392-020-0106-1
- Wang DT, Huang RH, Cheng X, Zhang ZH, Yang YJ, Lin X. Tanshinone IIA attenuates renal fibrosis and inflammation via altering expression of TGF- β /smad and NF- κ B signaling pathway in 5/6 nephrectomized rats. *Int Immunopharmacol.* (2015) 26(1):4–12. doi: 10.1016/j.intimp.2015.02.027
- Doke T, Abedini A, Aldridge DL, Yang YW, Park J, Hernandez CM, et al. Single-cell analysis identifies the interaction of altered renal tubules with basophils orchestrating kidney fibrosis. *Nat Immunol.* (2022) 23(6):947–59. doi: 10.1038/s41590-022-01200-7
- Brown AJ, Joseph PR, Sawant KV, Rajarathnam K. Chemokine CXCL7 heterodimers: structural insights, CXCR2 receptor function, and glycosaminoglycan interactions. *Int J Mol Sci.* (2017) 18(4):748. doi: 10.3390/ijms18040748
- Ahuja SK, Murphy PM. The CXC chemokines growth-regulated oncogene (GRO) α , GRO β , GRO γ , neutrophil-activating peptide-2, and epithelial cell-derived neutrophil-activating peptide-78 are potent agonists for the type B, but not the type A, human interleukin-8 receptor. *J Biol Chem.* (1996) 271(34):20545–50. doi: 10.1074/jbc.271.34.20545
- Page C, Pitchford S. Neutrophil and platelet complexes and their relevance to neutrophil recruitment and activation. *Int Immunopharmacol.* (2013) 17(4):1176–84. doi: 10.1016/j.intimp.2013.06.004
- Sreeramkumar V, Adrover JM, Ballesteros I, Cuartero MI, Rossaint J, Bilbao I, et al. Neutrophils scan for activated platelets to initiate inflammation. *Science.* (2014) 346(6214):1234–8. doi: 10.1126/science.1256478
- Mangold A, Alias S, Scherz T, Hofbauer M, Jakowitsch J, Panzenböck A, et al. Coronary neutrophil extracellular trap burden and deoxyribonuclease activity in ST-elevation acute coronary syndrome are predictors of ST-segment resolution and infarct size. *Circ Res.* (2015) 116(7):1182–92. doi: 10.1161/CIRCRESAHA.116.304944
- Andrews RK, Arthur JF, Gardiner EE. Neutrophil extracellular traps (NETs) and the role of platelets in infection. *Thromb Haemostas.* (2014) 112(4):659–65. doi: 10.1160/TH14-05-0455
- Ghasemzadeh M, Kaplan ZS, Alwis I, Schoenwaelder SM, Ashworth KJ, Westein E, et al. The CXCR1/2 ligand NAP-2 promotes directed intravascular leukocyte migration through platelet thrombi. *Blood.* (2013) 121(22):4555–66. doi: 10.1182/blood-2012-09-459636
- Brown AJ, Sepuru KM, Sawant KV, Rajarathnam K. Platelet-derived chemokine CXCL7 dimer preferentially exists in the glycosaminoglycan-bound form: implications for neutrophil-platelet crosstalk. *Front Immunol.* (2017) 8:1248. doi: 10.3389/fimmu.2017.01248
- Hayderi A, Kumawat AK, Shavva VS, Dreifaldt M, Sigvant B, Petri MH, et al. RSAD2 is abundant in atherosclerotic plaques and promotes interferon-induced CXCR3-chemokines in human smooth muscle cells. *Sci Rep.* (2024) 14(1):8196. doi: 10.1038/s41598-024-58592-9
- Huber TB, Reinhardt HC, Exner M, Burger JA, Kerjaschki D, Saleem MA, et al. Expression of functional CCR and CXCR chemokine receptors in podocytes. *J Immunol.* (2002) 168(12):6244–52. doi: 10.4049/jimmunol.168.12.6244
- Acosta JC, O'Loghlen A, Banito A, Gujarró MV, Augert A, Raguz S, et al. Chemokine signaling via the CXCR2 receptor reinforces senescence. *Cell.* (2008) 133(6):1006–18. doi: 10.1016/j.cell.2008.03.038
- Shu W, Yang H, Zhang L, Lu MM, Morrissey EE. Characterization of a new subfamily of winged-helix/forkhead (fox) genes that are expressed in the lung and act as transcriptional repressors. *J Biol Chem.* (2001) 276(29):27488–97. doi: 10.1074/jbc.M100636200
- Zhang SP, Yang RH, Shang J, Gao T, Wang R, Peng XD, et al. FOXC1 up-regulates the expression of toll-like receptors in myocardial ischemia. *J Cell Mol Med.* (2019) 23(11):7566–80. doi: 10.1111/jcmm.14626
- Liu XM, Du SL, Miao R, Wang LF, Zhong JC. Targeting the forkhead box protein P1 pathway as a novel therapeutic approach for cardiovascular diseases. *Heart Fail Rev.* (2022) 27(1):345–55. doi: 10.1007/s10741-020-09992-2
- Li B, Xiong W, Zuo W, Shi Y, Wang T, Chang L, et al. Proximal telomeric decompaction due to telomere shortening drives FOXC1-dependent myocardial senescence. *Nucleic Acids Res.* (2024) 52(11):6269–84. doi: 10.1093/nar/gkae274
- Igarashi W, Takagi D, Okada D, Kobayashi D, Oka M, Io T, et al. Bioinformatic identification of potential RNA alterations on the atrial fibrillation remodeling from human pulmonary veins. *Int J Mol Sci.* (2023) 24(13):10501. doi: 10.3390/ijms241310501

44. Miyashita N, Horie M, Suzuki HI, Saito M, Mikami Y, Okuda K, et al. FOXL1 regulates lung fibroblast function via multiple mechanisms. *Am J Respir Cell Mol Biol*. (2020) 63(6):831–42. doi: 10.1165/rcmb.2019-0396OC
45. Zhang L, Tian M, Zhang M, Li C, Wang X, Long Y, et al. Forkhead box protein K1 promotes chronic kidney disease by driving glycolysis in tubular epithelial cells. *Adv Sci (Weinh)*. (2024) 11(36):e2405325. doi: 10.1002/advs.202405325
46. Scheenstra MR, Salunkhe V, De Cuyper IM, Hoogenboezem M, Li E, Kuijpers TW, et al. Characterization of hematopoietic GATA transcription factor expression in mouse and human dendritic cells. *Blood Cells Mol Dis*. (2015) 55(4):293–303. doi: 10.1016/j.bcmd.2015.07.006
47. Yang X, Zhang Q, Li S, Devarajan R, Luo B, Tan Z, et al. GATA2 co-opts TGFβ1/SMAD4 oncogenic signaling and inherited variants at 6q22 to modulate prostate cancer progression. *J Exp Clin Cancer Res*. (2023) 42(1):198. doi: 10.1186/s13046-023-02745-7
48. Broman MT, Nadadur RD, Perez-Cervantes C, Burnicka-Turek O, Lazarevic S, Gams A, et al. A genomic link from heart failure to atrial fibrillation risk: FOG2 modulates a TBX5/GATA4-dependent atrial gene regulatory network. *Circulation*. (2024) 149(15):1205–30. doi: 10.1161/CIRCULATIONAHA.123.066804
49. Yin C, Vrieze AM, Rosoga M, Akingbasote J, Pawlak EN, Jacob RA, et al. Efferocytic defects in early atherosclerosis are driven by GATA2 overexpression in macrophages. *Front Immunol*. (2020) 11:594136. doi: 10.3389/fimmu.2020.594136
50. Wu TT, Tai YT, Cherg YG, Chen TG, Lin CJ, Chen TL, et al. GATA-2 transduces LPS-induced il-1β gene expression in macrophages via a toll-like receptor 4/MD88/MAPK-dependent mechanism. *PLoS One*. (2013) 8(8):e72404. doi: 10.1371/journal.pone.0072404
51. Zhang Y, Huang H, Kong Y, Xu C, Dai L, Geng X, et al. Kidney tubular transcription co-activator, yes-associated protein 1 (YAP), controls the expression of collecting duct aquaporins and water homeostasis. *Kidney Int*. (2023) 103(3):501–13. doi: 10.1016/j.kint.2022.10.007
52. Yu L, Moriguchi T, Kaneko H, Hayashi M, Hasegawa A, Nezu M, et al. Reducing inflammatory cytokine production from renal collecting duct cells by inhibiting GATA2 ameliorates acute kidney injury. *Mol Cell Biol*. (2017) 37(22):e00211-17. doi: 10.1128/mcb.00211-17
53. Nesapiragasan V, Hayiroglu M, Sciacca V, Sommer P, Sohns C, Fink T. Catheter ablation approaches for the treatment of arrhythmia recurrence in patients with a durable pulmonary vein isolation. *Balkan Med J*. (2023) 40(6):386–94. doi: 10.4274/balkanmedj.galenos.2023.2023-9-48
54. Wang X, Zhao Q, Huang H, Tang Y, Xiao J, Dai Z, et al. Effect of renal sympathetic denervation on atrial substrate remodeling in ambulatory canines with prolonged atrial pacing. *PLoS One*. (2013) 8(5):e64611. doi: 10.1371/journal.pone.0064611
55. Steinberg JS, Shabanov V, Ponomarev D, Losik D, Ivanickiy E, Kropotkin E, et al. Effect of renal denervation and catheter ablation vs. catheter ablation alone on atrial fibrillation recurrence among patients with paroxysmal atrial fibrillation and hypertension: the ERADICATE-AF randomized clinical trial. *JAMA*. (2020) 323(3):248–55. doi: 10.1001/jama.2019.21187
56. Cluett JL, Blazek O, Brown AL, East C, Ferdinand KC, Fisher NDL, et al. Renal denervation for the treatment of hypertension: a scientific statement from the American Heart Association. *Hypertension*. (2024) 81(10):e135–48. doi: 10.1161/hyp.0000000000000240
57. Akumwami S, Morishita A, Iradukunda A, Kobara H, Nishiyama A. Possible organ-protective effects of renal denervation: insights from basic studies. *Hypertens Res*. (2023) 46(12):2661–9. doi: 10.1038/s41440-023-01393-w



OPEN ACCESS

EDITED BY

Tim Hulsen,
Rotterdam University of Applied Sciences,
Netherlands

REVIEWED BY

Natallia Maroz-Vadalazhskaya,
Belarusian State Medical University, Belarus
Jafar A. Alzubi,
Al-Balqa Applied University, Jordan
Michael Guckert,
Technische Hochschule Mittelhessen,
Germany

*CORRESPONDENCE

Hsin-Yueh Liang
✉ liangsy2@gmail.com

RECEIVED 13 September 2024

ACCEPTED 27 February 2025

PUBLISHED 17 March 2025

CITATION

Liang H-Y, Hsu K-C, Chien S-Y, Yeh C-Y, Sun T-H, Liu M-H and Ng KK (2025) Deep learning analysis of exercise stress electrocardiography for identification of significant coronary artery disease.

Front. Artif. Intell. 8:1496109.

doi: 10.3389/frai.2025.1496109

COPYRIGHT

© 2025 Liang, Hsu, Chien, Yeh, Sun, Liu and Ng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning analysis of exercise stress electrocardiography for identification of significant coronary artery disease

Hsin-Yueh Liang^{1,2*}, Kai-Cheng Hsu^{3,4,5}, Shang-Yu Chien³, Chen-Yu Yeh³, Ting-Hsuan Sun³, Meng-Hsuan Liu³ and Kee Koon Ng¹

¹Division of Cardiology, Department of Medicine, China Medical University Hospital, Taichung, Taiwan,

²Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, Taiwan, ³Artificial Intelligence Center, China Medical University Hospital, Taichung, Taiwan, ⁴School of Medicine, China Medical University, Taichung, Taiwan, ⁵Department of Neurology, China Medical University Hospital, Taichung, Taiwan

Background: The diagnostic power of exercise stress electrocardiography (ExECG) remains limited. We aimed to construct an artificial intelligence (AI)-based method to enhance ExECG performance to identify patients with significant coronary artery disease (CAD).

Methods: We retrospectively collected 818 patients who underwent both ExECG and coronary angiography (CAG) within 6 months. The mean age was 57.0 ± 10.1 years, and 614 (75%) were male patients. Significant coronary artery disease was seen in 369 (43.8%) CAG reports. We also included 197 individuals with normal ExECG and low risk of CAD. A convolutional recurrent neural network algorithm, integrating electrocardiographic (ECG) signals and features from ExECG reports, was developed to predict the risk of significant CAD. We also investigated the optimal number of inputted ECG signal slices and features and the weighting of features for model performance.

Results: Using the data of patients undergoing CAG for training and test sets, our algorithm had an area under the curve, sensitivity, and specificity of 0.74, 0.86, and 0.47, respectively, which increased to 0.83, 0.89, and 0.60, respectively, after enrolling 197 subjects with low risk of CAD. Three ECG signal slices and 12 features yielded optimal performance metrics. The principal predictive feature variables were sex, maximum heart rate, and ST/HR index. Our model generated results within one minute after completing ExECG.

Conclusion: The multimodal AI algorithm, leveraging deep learning techniques, efficiently and accurately identifies patients with significant CAD using ExECG data, aiding clinical screening in both symptomatic and asymptomatic patients. Nevertheless, the specificity remains moderate (0.60), suggesting a potential for false positives and highlighting the need for further investigation.

KEYWORDS

exercise stress electrocardiography, coronary artery disease, deep learning, multimodal approach, feature variable, artificial intelligence, clinical screening, convolutional recurrent neural network

Introduction

Ischemic heart disease is the major cause of mortality worldwide. Recent findings from the global disease burden study indicate that ischemic heart disease caused more than 9 million deaths in 2021 (Vaduganathan et al., 2022; Malakar et al., 2019). Early diagnosis is crucial because lifestyle modification and medical intervention improve life quality and prolong survival (Knuuti et al., 2020).

The workup of a patient presented with suspected coronary artery disease (CAD) involves history taking, physical examination, and initial examinations. Possible CAD is further evaluated using many noninvasive test modalities, including exercise stress electrocardiography (ExECG), stress echocardiography, stress nuclear myocardial perfusion imaging, cardiovascular magnetic resonance imaging, and coronary computed tomography angiography (CCTA). Among them, ExECG, which has been used for >60 years, is a safe and affordable test for suspected CAD. However, although several ExECG scores, such as the Duke treadmill score, have been developed to improve diagnostic accuracy, the diagnostic power of ExECG remains limited with an area under the receiver operating characteristics curve (AUC) of 0.72–0.76 (Shaw et al., 1998).

Artificial intelligence (AI) has been applied in many disease models (Yadav and Jadhav, 2019; Movassagh et al., 2023; Alzubi et al., 2021; Kose et al., 2021). Given the limitations in the diagnostic accuracy of ExECG, AI offers the potential to overcome these challenges by detecting subtle patterns in ExECG data that might be missed by conventional interpretation methods. AI-enabled ExECG algorithms, which utilize various models and datasets to enhance accuracy, efficiency, and applicability of CAD prediction, have been published, with AUC, sensitivity, and specificity of 0.73–0.78, 0.25–0.85, and 0.43–0.97, respectively (Babaoglu et al., 2009; Babaoglu et al., 2010; Yilmaz et al., 2023; Lee et al., 2022). A hybrid convolutional neural network (CNN)–long short-term memory (LSTM) architecture has been shown to effectively process and analyze electrocardiography (ECG) (Banerjee et al., 2020; Cheng et al., 2021). We hypothesized that the application of a hybrid CNN–LSTM model in ExECG might accurately and efficiently identify patients with significant CAD. To test this hypothesis, we conducted a retrospective study of ExECG in patients who underwent invasive coronary angiography (CAG) and those with normal ExECG to develop and validate a deep learning AI model to predict significant CAD.

The primary objectives of this study were to: (1) develop a hybrid CNN–LSTM algorithm that integrates ECG signals and ExECG features to improve the diagnostic accuracy of ExECG for significant CAD; (2) optimize the number and weighting of ECG signal segments and features to maximize model performance; and (3) evaluate the model's efficiency in generating results. Our findings suggest that a multimodal AI algorithm leveraging deep learning can rapidly and accurately detect significant CAD from ExECG data, delivering results within one minute post-test. This advancement holds potential to enhance clinical screening for CAD in both symptomatic and asymptomatic patients.

Materials and methods

Study population

We enrolled 4,959 ExECG reports of 4,849 patients who underwent symptom-limited ExECG saved in XML format using the GE CASE 6.73 Stress Test system from January 2017 to January 2022 (Fletcher et al., 2013). We excluded patients with incomplete ExECG data or pacemaker implantation. The CAG group was defined as patients who underwent ExECG and subsequent CAG within 6 months, which was further divided into two subgroups with (A) and without (N) significant CAD. ExECG reports showing peak heart rates >85% of the maximum predicted rate and interpreted as normal by cardiologists, without subsequent CAG within 6 months, were categorized as subgroup T. In this group, patients with known or suspected CAD, hypertension, hyperlipidemia, diabetes, or clinical risk factors of CAD [male ≥ 45 years old, female ≥ 55 years old, or body mass index ≥ 24 or < 18.5 kg/m²] were further excluded, resulting in a subgroup at low risk of CAD (H). We selected patients who were evaluated through both ExECG and CCTA within 6 months at the Health Screening Center and had <50% coronary artery stenosis and classified them into subgroup C (Figure 1). Due to the variable positive predictive value of CCTA, ranging from 64 to 91%, patients identified by CCTA as having >50% coronary artery stenosis were excluded from the study (Arbab-Zadeh and Hoe, 2011). We included patients with a wide range of CAD severity to ensure the generalizability of our model.

We conducted model training using combinations of the aforementioned subgroups, resulting in five groups, including the CAG group (subgroups N and A); group II (subgroups N, A, and T); group III (subgroups N, A, and H); group IV (subgroups N, A, and C); and group V (subgroups N, A, H, and C). Some patients underwent more than one examination. We used any complete examination reports available, which resulted in a discrepancy between the number of examinations and patients.

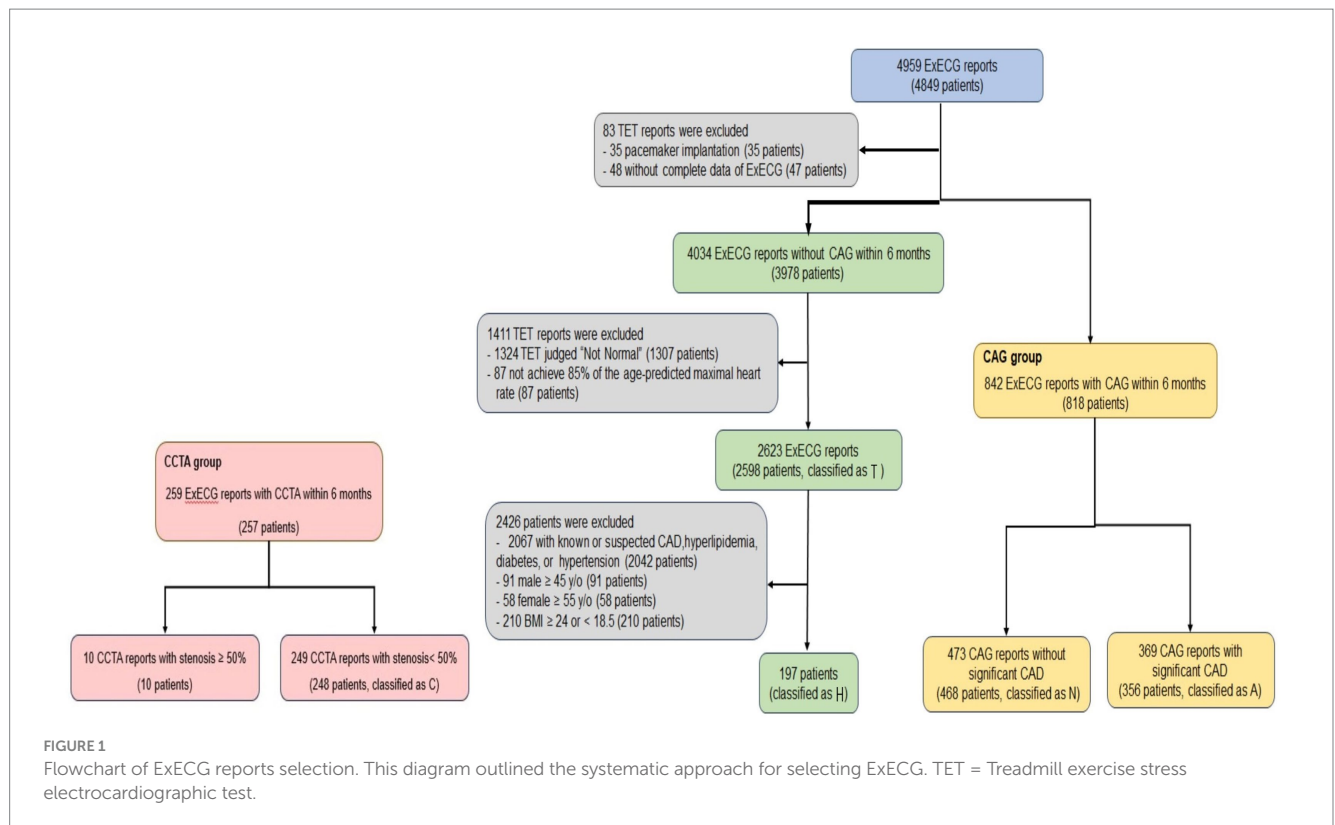
Patients who underwent ExECG and CAG within 6 months at our institute and Asia University Hospital after February 2022 were used for external validation, respectively. The Institutional Review Board (IRB) of China Medical University Hospital approved this retrospective, single-center study (IRB number CMUH110-REC3-019).

Coronary angiography

Coronary angiography was performed using standard techniques and used as the golden standard to determine the presence and severity of CAD. Significant CAD was defined as $\geq 50\%$ stenosis of the left main stem and/or $\geq 70\%$ stenosis in any major coronary artery (Nallamothu et al., 2013; Bhatt, 2015).

ExECG data retrieval

The ExECG equipment generates a 10-s ECG signal at a frequency of 500 Hz in 12 ECG leads. ECGs in the pretest, exercise (peak heart rate), and recovery phases were retrieved for model 1



training (Figure 2, left panel). To investigate if the number of ECG signal slices for model input is proportional to the model's capability, we added three extra slices close to the peak heart rate during the exercise and recovery phases for model 2 training (Figure 2, right panel).

The ExECG reports included supplementary physiological data alongside the ECG signals, which have been shown to enhance the accuracy of ExECG assessments (Lehtinen, 1999; Christman et al., 2014; Ahmed et al., 2015; Marzlin and Webner, 2019; Schultz et al., 2017; Ghaffari et al., 2017; Mieres et al., 2014; Snader et al., 1997). To improve the model's effectiveness in detecting CAD, we integrated a carefully selected set of these physiological metrics as metadata, consisting of 14 primary features and two derived features. The primary features included sex, age, BMI, resting and peak heart rates, maximum predicted heart rate, resting and peak systolic and diastolic blood pressures, maximum rate-pressure product, maximum workload, maximum ST depression, and the ST/heart rate index. Additionally, we derived chronotropic incompetence and percent predicted metabolic equivalents to further support the model's predictive capability (Supplementary material).

ECG signals and metadata preprocessing

Our model processed input data as stacked ECG signals, annotated as $5,000 \times 12$ in three phases (pretest, exercise, and recovery), indicating that each lead had 5,000 data points. Subsequently, the ECG signals were separated into limb (I–III, aVR, aVL, and aVF) and precordial leads (V1–V6). Each lead, composed of one-dimensional data, was directed into an analytical module,

structure A, for subsequent processing (Figure 3) (Yildirim et al., 2018; Kiranyaz et al., 2015).

Numerical and categorical variables of metadata were preprocessed using MinMaxScaler and one-hot encoder techniques. We excluded the ExECG if sex was absent. In our study, the missing data rate was minimal (0–0.5%), and the data followed a normal or near-normal distribution. For other variables with incomplete entries, we used the mean of the data, ensuring the continuity and integrity of the dataset for analysis. Outliers were individually reviewed to ensure accuracy. If it was not feasible to confirm whether an outlier was accurate, it was handled using the same approach as missing data.

Model design

For our model development, we designed a deep learning framework that integrated the CNN with LSTM network, as previously described (Chen et al., 2022) (Figure 3). Structure A in the architecture extracted the characteristics of ECG signals from the limb and precordial leads in each phase, and it consisted of six layers of one-dimensional CNN layers (Figure 4). We added a dropout layer after every three CNN layers, randomly discarding 20% of the information, to prevent overfitting. The Leaky Rectified Linear Unit (LeakyReLU) activation function was also used for each layer in structure A, which maintained the gradient flow during the training process, potentially leading to a better model performance. The output results from structure A were combined and input into the attention layer (Yang et al., 2016) to determine important weight vectors, followed by two bidirectional LSTM layers for sequence analysis. Subsequently, the output data were entered into two dense layers to classify the processed characteristics of ECG signals (Figure 3).

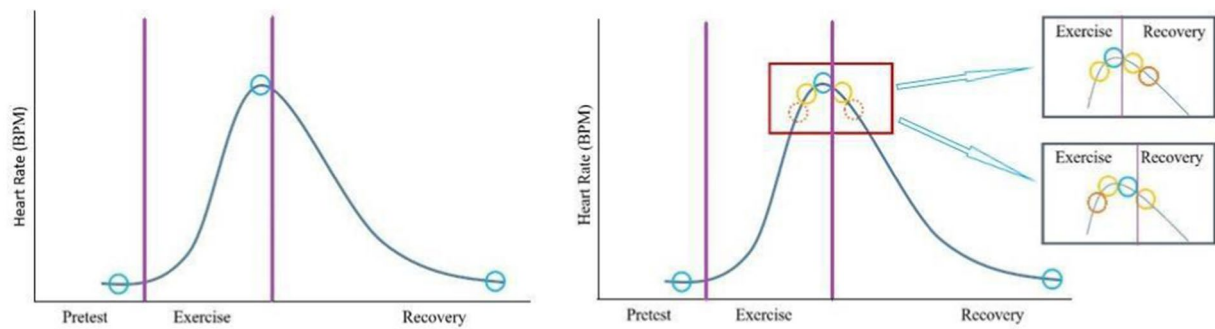


FIGURE 2
Selection of electrocardiographic slices. In model 1 (left panel), one slice (blue circle) of the signal in each stage was selected for training. In model 2 (right panel), three extra slices (yellow and solid orange circles) close to the peak heart rates during exercise and recovery phases were added.

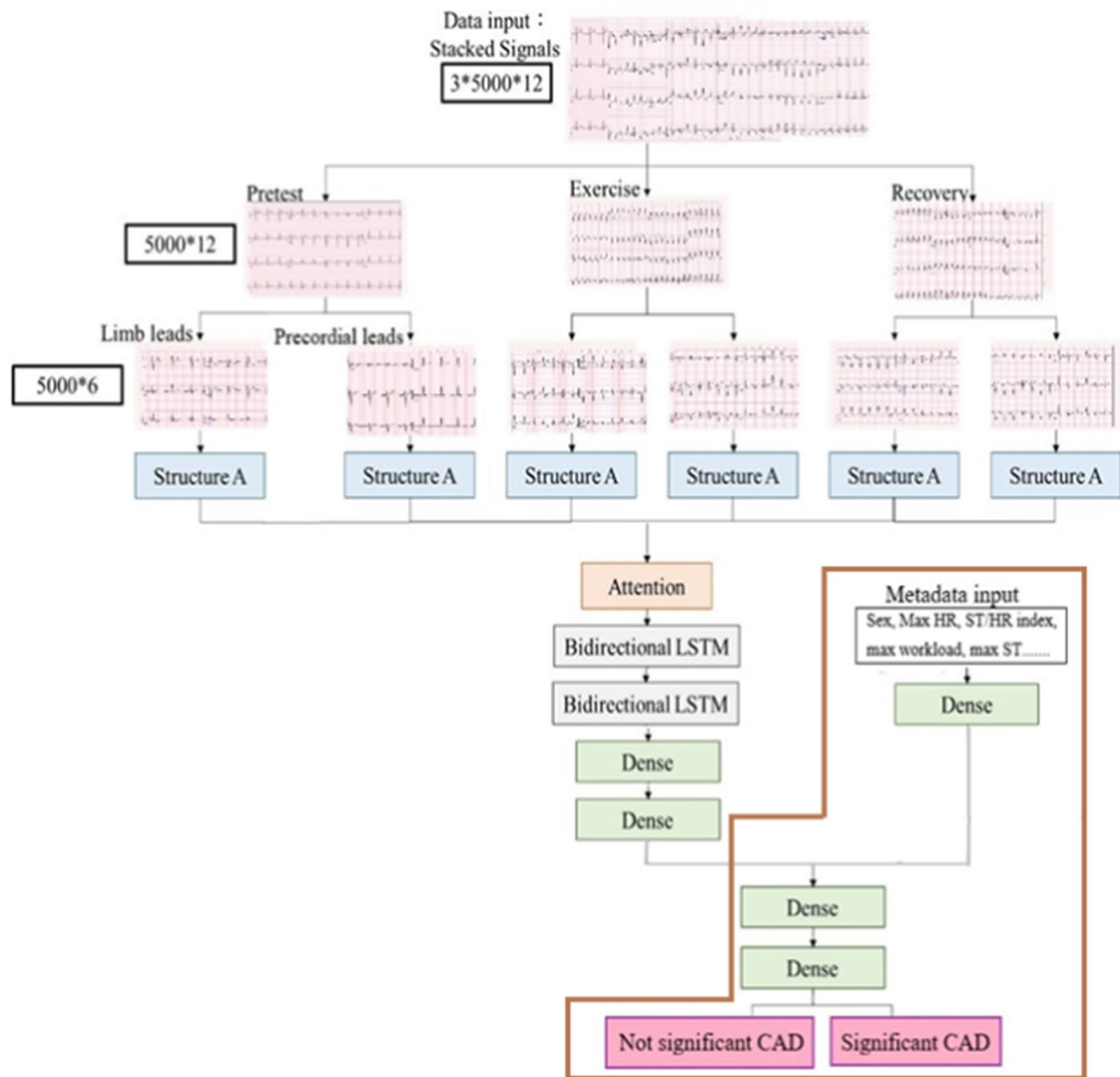


FIGURE 3
Process of ExECG analysis. It comprised data acquisition, preprocessing, feature extraction, and machine learning.

Quantity of ECG slice and features

In the CAG group (A + N), optimal performance was achieved by integrating three ECG slices (pretest, peak heart rate, and recovery phases) and 12 features without blood pressure data, resulting in AUC, sensitivity, and specificity of 0.74, 0.86, and 0.47, respectively (Table 2). More slices and/or features did not improve model performance. The SHapley Additive exPlanations summary plot (Figure 5) showed that sex, maximum HR, and ST/HR index were the most significant predictors in our model. Therefore, we used the integration of three ECG slices and 12 features in the next stage.

Performance across five training groups

Subsequently, the model was trained using various group combinations. The outcomes across five training groups were not significantly different if the test set only included the CAG group (A + N) (Table 3), with AUC, sensitivity, and specificity of 0.74–0.78, 0.82–0.89, and 0.47–0.51, respectively. When T subgroup was integrated into the training and test sets (group II), accuracy significantly improved, with AUC, NPV, and specificity of 0.75, 0.88, 0.97, and 0.75, respectively. However, this enhancement was achieved with detriment to the F1, and PPV. When low-risk patient data (H) were included for both training and testing (group III), accuracy significantly improved, with AUC, NPV, and specificity of 0.71, 0.83, 0.90, and 0.60, respectively. However, the F1 score, PPV, and sensitivity had minimal variation. The demographic, CAG, and ExECG features did not differ significantly between training, validation, and testing sets in group III (Supplementary Table 1). The performance was not significantly different when the model was trained and tested on groups III, IV, or V (Table 3). We also compared our models with the conventional ExECG algorithm, which primarily focuses on exercise-induced ST-segment changes as assessed by board-certified cardiologists with varying levels of clinical experience (Mieres et al., 2014). Compared with the conventional ExECG algorithm, the performance of our AI model showed improvements across all measured variables and delivered predictive outcomes within 1 min after completing ExECG.

TABLE 2 Model performance across varied quantities of inputted ECGs and features.

Inputted number		Model performance				
ECGs	Features	ACC	AUC	F1	SEN	SPE
N = 6	0	0.62	0.66	0.58	0.58	0.64
	10	0.64	0.66	0.59	0.57	0.69
	12	0.64	0.69	0.68	0.83	0.47
	14	0.65	0.70	0.68	0.82	0.50
	16	0.68	0.72	0.68	0.75	0.61
N = 3	0	0.65	0.68	0.66	0.72	0.59
	10	0.60	0.68	0.65	0.79	0.44
	12	0.65	0.74	0.70	0.86	0.47
	14	0.66	0.72	0.64	0.65	0.68
	16	0.64	0.71	0.68	0.82	0.48

ACC = accuracy; AUC = area under the curve; F1 = F1 score; SEN = sensitivity; SPE = specificity. Bold value indicated optimal performance.

The performance of the model using the combination of ECG signals and metadata was comparable to that using metadata alone (the brown box in Figure 3), as shown in Table 4. However, the model integrating ECG signals and metadata demonstrated higher sensitivity compared to the model relying solely on metadata.

External validation

The comparative analysis of performance across five training groups showed minimal discrepancy in patients at our institute after February 2022 and Asia University Hospital (Table 3), indicating that our models can potentially be applied in diverse settings.

Bootstrap validation

We implemented bootstrapping to assess low AUC probability of Table 2 and simulated the sample size improvement ($n = 10,000$) with a flexible bootstrap distribution. The results showed that the AUC distribution of our model was stable, with minimal variability and a narrow confidence interval. The consistency between bootstrap estimates and the original value underscored the reliability of this method for performance validation using SAS JMP Academic Suite Version 17.2 (JMP Inc., NC, United States) (Figure 6).

Discussion

Our study provides an efficient and accurate tool to identify patients with significant CAD by the AI-enhanced ExECG algorithm, which achieved an AUC of 0.83, a sensitivity of 0.89, and a specificity of 0.60, within 1 min. The most important feature predictors for our model performance were sex, maximum heart rate, and ST/HR index.

The conventional ExECG algorithm mainly depends on ST segment changes (Gibbons et al., 2002). A meta-analysis of 147 studies involving 24,047 patients reported mean sensitivity and specificity of 68 and 77%, respectively, but with considerable variability, ranging from 23–100% for sensitivity and 17–100% for specificity. The variation in diagnostic accuracy could be attributed to significant disparities in the demographic and clinical profiles of the studied cohorts, divergent criteria for defining the presence and severity of CAD, and differences in the selection of diagnostic variables (Fletcher et al., 2013; Detrano et al., 1989). Compared with the conventional ExECG algorithm, our model, incorporating ECG signal along with 12 features, showed superior performance.

In 2009, Babaoglu et al. initially explored the use of AI algorithms to detect and localize CAD through ExECG (Babaoglu et al., 2009). Their methodology incorporated 27 distinct features as inputs into their model. Subsequently, they refined their approach by reducing the feature set to 18 and applied the support vector machine method for further studies (Babaoglu et al., 2010). Various models have been developed with the rapid evolution of machine learning technologies. Lee et al. introduced the random forest algorithm to enhance ExECG diagnostic capabilities, utilizing a dataset comprising 30 specific features, with the option to incorporate clinical data (Lee et al., 2022). Following this advancement, Yilmaz et al. implemented the eXtreme gradient boosting algorithm, capitalizing on ECG characteristics and signals presented in JPEG format for their analysis (Yilmaz et al., 2023). Compared with a previous study, our training model using five different patient groups showed not inferior performance, supported by AUC metrics, when assessed in the CAG group (Table 5). The AUC values reported in these

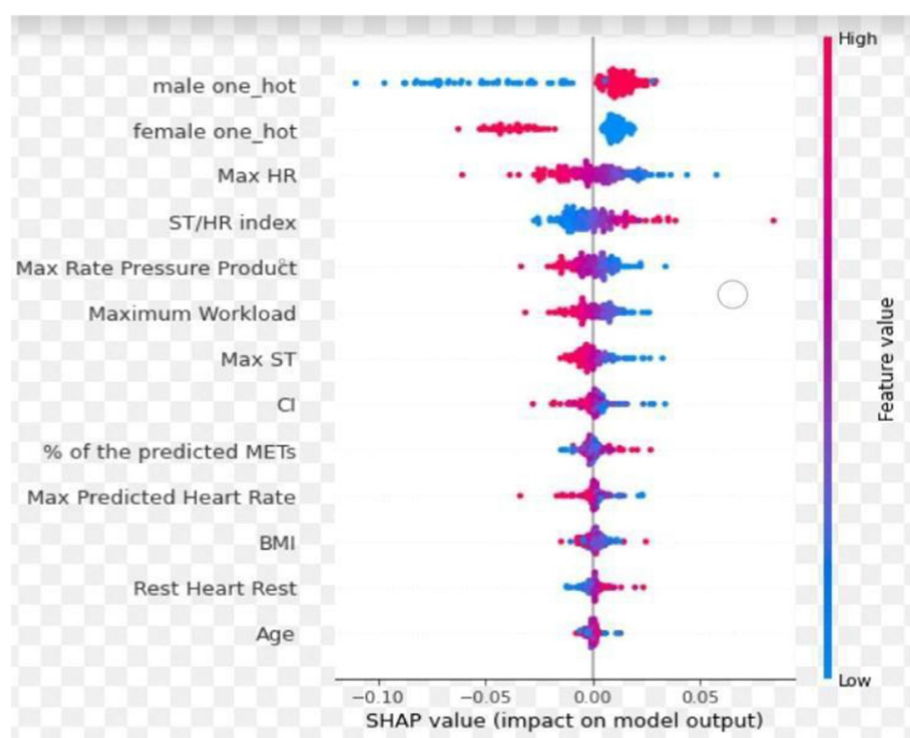


FIGURE 5
The SHapley additive exPlanations summary plot. The red and blue colors represent the feature values, with red indicating high feature values and blue corresponding to low feature values. Positive SHAP values indicate that the feature increases the likelihood of predicting significant CAD, while negative SHAP values suggest a decreased likelihood of significant CAD. Sex, maximum HR and ST/HR index were the most significant predictors in our model.

TABLE 3 Performance comparison of the model integrating ECG signals and metadata across subgroup combinations.

Training set	Test set	ACC	AUC	F1	PPV	NPV	SEN	SPE
Gr CAG, N: A	CAG, N: A	0.65	0.74	0.70	0.59	0.79	0.86	0.47
	N: A*	0.63	0.74	0.69	0.58	0.73	0.84	0.42
	N: A#	0.55	0.70	0.62	0.49	0.73	0.84	0.34
Gr II, (T + N): A	CAG, N: A	0.66	0.76	0.69	0.60	0.77	0.82	0.51
	N: A*	0.59	0.73	0.66	0.56	0.69	0.83	0.37
	N: A#	0.61	0.68	0.66	0.53	0.82	0.88	0.42
	II, (T + N): A	0.75	0.88	0.43	0.29	0.97	0.82	0.75
Gr III, (H + N): A	CAG, N: A	0.66	0.78	0.71	0.59	0.82	0.89	0.47
	N: A*	0.61	0.75	0.69	0.57	0.77	0.89	0.35
	N: A#	0.55	0.63	0.63	0.49	0.77	0.88	0.31
	III, (H + N): A	0.71	0.83	0.69	0.57	0.90	0.89	0.60
Gr IV, (C + N): A	CAG, N: A	0.67	0.78	0.71	0.60	0.83	0.89	0.48
	N: A*	0.62	0.75	0.69	0.57	0.78	0.89	0.35
	N: A#	0.56	0.63	0.63	0.49	0.78	0.88	0.32
	IV, (C + N): A	0.65	0.79	0.65	0.51	0.89	0.89	0.52
Gr V, (C + H + N): A	CAG, N: A	0.66	0.78	0.71	0.59	0.82	0.89	0.47
	N: A*	0.62	0.75	0.69	0.57	0.78	0.89	0.35
	N: A#	0.56	0.63	0.63	0.49	0.78	0.88	0.32
	V, (C + H + N): A	0.68	0.82	0.63	0.49	0.92	0.89	0.59
Conventional	CAG, N: A	0.46	-	0.55	0.45	0.49	0.71	0.24

* External validation from data of our institute after Feb 2022. # External validation from data of Asia university hospital. ACC = accuracy; AUC = area under the ROC curve; F1 = F1 score; PPV = positive predictive value; NPV = negative predictive value; SEN = sensitivity; SPE = specificity.

TABLE 4 Comparison of model performance using ECG signals and metadata versus metadata alone.

Patients	Model	ACC	AUC	F1	PPV	NPV	SEN	SPE
Gr CAG, N: A	ECG+ metadata	0.65	0.74	0.70	0.59	0.79	0.86	0.47
	Only metadata	0.66	0.72	0.67	0.61	0.72	0.73	0.59

ExECG-CAG studies were not as high as those observed in other AI implementations for ECG analyses, such as those for arrhythmia and systolic dysfunction (Attia et al., 2019; Adedinsawo et al., 2020). The selection bias inherent in ExECG-CAG studies might account for this discrepancy. Specifically, patients selected for CAG typically showed a higher probability of having obstructive CAD, a selection criterion that usually excludes healthy individuals. Consequently, this predisposition influenced the severity spectrum used during model training, culminating in acceptable but not outstanding AUC values.

Healthy individuals are commonly included into the training sets to enhance the generalization capability of AI models, facilitating its broader applicability across patients at various risks (Huang et al., 2022). Thus, we included individuals with normal ExECG interpreted by cardiologists and those exhibiting insignificant coronary artery stenosis determined by CCTA into the training sets (groups II–V) (Table 3). The performance metrics for each group (groups II–V) significantly improved, with AUC, sensitivity, and specificity of 0.79–0.88, 0.82–0.89, and 0.52–0.75, respectively. Although accuracy and AUC were the highest when the model underwent both training and testing on group II, PPV and F1 score significantly decreased, with increased specificity. This observation suggests a propensity for the model, when trained with data from group II, to exhibit a bias towards classifying subjects as normal. This alteration may be attributed to data imbalance, considering that the sample size of subgroup T disproportionately exceeded that of subgroup A (Huang et al., 2023). Furthermore, the possibility of silent ischemia in subgroup T could not be entirely ruled out, potentially contributing to a reduction in both PPV and F1 score. Conversely, the outcomes of the model trained and tested on group III including patients at low risk (H) also showed excellent discrimination, supported by an AUC, sensitivity, and specificity of 0.83, 0.89, and 0.60, respectively, which was achieved without compromising the PPV and F1 score. Considering that coronary artery disease is treatable yet often presents acutely and can lead to severe complications, we prioritized designing an algorithm with greater sensitivity and accuracy to assist in the early identification of potential cases. The performance of the model when using group IV or V did not surpass that noted in group III. This observation may be due to the patients undergoing CCTA (included within groups IV or V) are generally older and possess higher cardiovascular disease risk factors, aligning more closely with the characteristics of subgroup A instead of subgroup H (Table 1). These demographic and clinical characteristics may potentially lead to model misinterpretation. Similarly, although the performance of the model using only metadata was comparable to that of the model integrating ECG signals and metadata (Table 4), we prioritized higher sensitivity, which led to our selection of the model combining ECG signals and metadata.

The ECG images presented in PDF or JPEG formats in previous studies underwent processing by the equipment and limited the display of each lead to 2.5 s. In contrast, we used original ECG signals directly generated by the equipment, extending the duration for each lead to 10 s, which enhanced the model's access to comprehensive ECG data. By

combining CNNs and LSTMs into a CRNN architecture, our model provides the benefit of both spatial feature extraction and temporal sequence modeling, allowing our model to understand the complex structure of ECG data, recognizing the immediate patterns in the signals and how these patterns change over time (Verma and Agarwal, 2018; Zhang et al., 2020; Zihlmann et al., 2017). In our investigation, the analysis of three ECG signal slices with 12 specific features during the pretest, peak heart rate, and recovery phase yielded optimal performance metrics. However, the additional ECG slices or features did not enhance predictive outcomes. The principal predictive variables were sex, maximum heart rate, and ST/HR index (Figure 5), offering valuable insights into the weighting of features to identify significant CAD (Lehtinen, 1999; Christman et al., 2014; Ahmed et al., 2015; Marzlin and Webner, 2019; Schultz et al., 2017; Ghaffari et al., 2017; Mieres et al., 2014; Snader et al., 1997). Moreover, our model can generate results within 1 min after completing ExECG. Future research should aim to enhance specificity by integrating clinical and imaging data, optimizing the AI algorithm, applying differential weighting during training, incorporating additional physiological features from the ExECG report, exploring the impact and significance of metadata, and expanding training datasets to include larger and more diverse populations (Benkarim et al., 2022; McKinney et al., 2020; Sato et al., 2022; Marwick et al., 1995; Gencbay et al., 1999; Siegler et al., 2011).

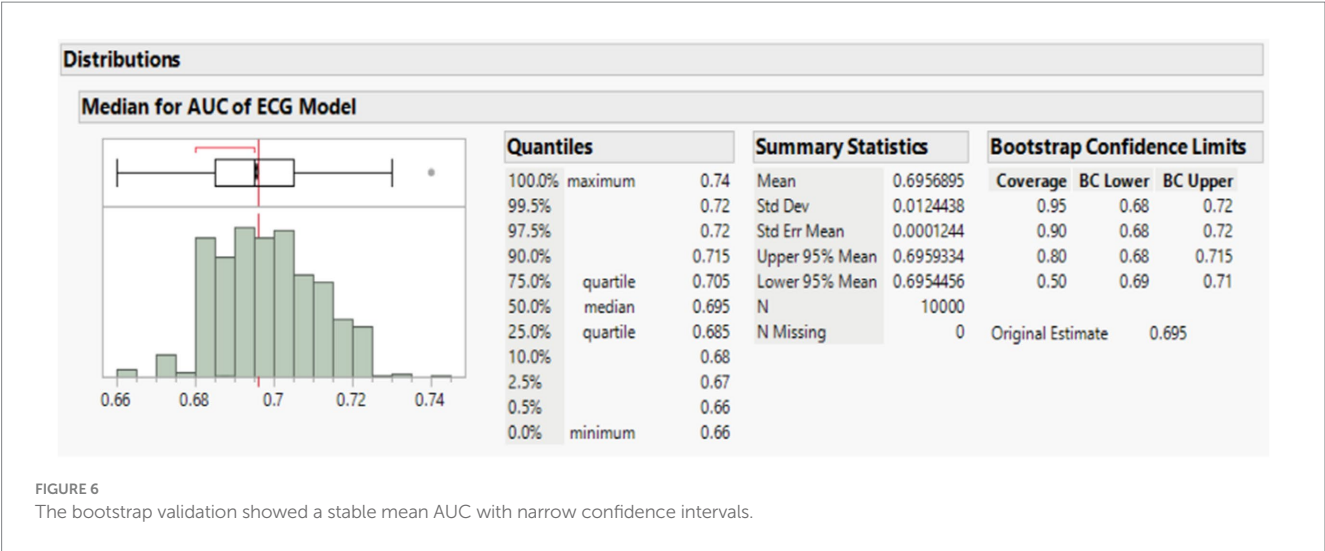
Limitations

Our study has limitations. First, the patient cohort was recruited from a single institution; notwithstanding, external validation was conducted at Asia University Hospital, revealing minimal variance in performance outcomes across the two facilities. Second, angiographic analyses were conducted by interventional cardiologists engaged in routine clinical practice rather than by a dedicated core laboratory, which may introduce a degree of subjective bias in interpretation. Despite these, significant stenosis was accurately identified consistently. Third, not all participants underwent angiography or CCTA; however, the likelihood of erroneously classifying patients with significant CAD as normal was reduced by preferentially selecting individuals at low risk for CAD. Additionally, our study focused on epicardial stenosis and did not assess coronary microvascular dysfunction. In some patients with diabetes or hypertension, coronary artery disease may arise from microvascular dysfunction rather than macrovascular stenosis (Vrints et al., 2024; O'Neal et al., 2017; Angeja et al., 2002; Okin et al., 2004). Using CAG-detected epicardial stenosis as the gold standard for evaluating functional tests like ExECG may not fully reflect the underlying pathophysiology and could negatively affect AI performance. Finally, low PPV and specificity might increase the incidence of unnecessary CAG. This limitation is primarily attributable to the selection bias inherent in ExECG-CAG studies. Moreover, our primary objective is to assist physicians in efficiently screening patients following ExECG, rather than acting as the sole determinant for advanced invasive

TABLE 5 The comparison with the existing literature.

Study	Cohort and definition of CAD	Input	Model	Output	ACC	AUC	SEN	SPE
Babaoglu et al. (2009)	Training and test cohorts 330 patients with ExECG and CAG within one month Definition of CAD Narrowing $\geq 50\%$ in LM, or narrowing $\geq 70\%$ in the other major coronary arteries	27 features from the ExECG data	ANN	LM	0.91	-	0.25	0.97
				LAD	0.73	-	0.75	0.70
				LCX	0.65	-	0.50	0.75
				RCA	0.69	-	0.50	0.80
Babaoglu et al. (2010)	Training and test cohorts 480 patients with ExECG and CAG within a month Definition of CAD Narrowing $\geq 50\%$ in LM, or narrowing $\geq 70\%$ in the other major coronary arteries	18 features from the ExECG data using the PCA method	SVM	CAD	0.79	-	-	-
Lee et al. (2022)	Training and test cohorts 2,325 patients with ExECG and CAG within a year Definition of CAD narrowing $\geq 70\%$ in LAD, LCX, RCA, or their main branches, or narrowing $\geq 50\%$ in LM	30 features from the ExECG data	RF	CAD	0.57	0.73	0.85	0.43
		30 features from the ExECG data and 7 clinical features		CAD	0.59	0.74	0.85	0.45
Yilmaz et al. (2023)	Training and test cohorts 294 patients with ExECG and CAG in the same month. Definition of CAD narrowing $\geq 70\%$ in LAD, LCX, RCA, or LM	23 extracted features from ExECG signal in JPEG format	XGBoost	CAD	0.81	0.78	0.67	0.85
Our study	Training and test cohorts 842 patients with ExECG and CAG within 6 months Definition of CAD Narrowing $\geq 50\%$ in LM, or narrowing $\geq 70\%$ in the other major coronary arteries	ExECG signals in XML format and 12 features	CRNN	CAD	0.65	0.74	0.86	0.47

CAD = coronary artery disease; ACC = accuracy; AUC = area under the ROC curve; SEN = sensitivity; SPE = specificity; ExECG = exercise electrocardiography; CAG = coronary angiography; LAD = left anterior descending artery; LCX = left circumflex artery; RCA = right coronary artery; LM = left main segment; XGBoost = eXtreme Gradient Boosting; RF = Random Forest; SVM = Support Vector Machine; ANN = Artificial Neural Network; CRNN = Convolutional Recurrent Neural Networks.



testing. Therefore, when AI-generated findings raise clinical uncertainty, additional imaging modalities should be considered. Further research aimed at improving specificity is warranted.

Conclusion

Our AI-based algorithm has shown promise in identifying patients with significant CAD using ExECG data. Integrating a multimodal approach that combines ECG signals with additional features enhances both predictive performance and efficiency. Further large-scale studies and algorithm refinements are needed to improve specificity and validate clinical utility across diverse patient populations.

Summary

This study aimed to develop an artificial intelligence (AI)-based method to enhance the efficiency and accuracy of exercise stress electrocardiography (ExECG) in detecting significant coronary artery disease (CAD). We retrospectively analyzed 818 patients who underwent both ExECG and coronary angiography (CAG) within 6 months. We used a Convolutional Recurrent Neural Network algorithm, which integrated electrocardiographic (ECG) signals and ExECG report features to predict significant CAD. The algorithm achieved an area under the curve (AUC) of 0.74, sensitivity of 0.86, and specificity of 0.47. With the inclusion of 197 low-risk patients, AUC, sensitivity, and specificity improved to 0.83, 0.89, and 0.60, respectively. Optimal performance was achieved with three ECG signal slices and 12 features, including sex, maximum heart rate, and ST/HR index as principal predictive variables. The AI model generated results within 1 min after completing ExECG, suggesting its potential to identify significant CAD efficiently and accurately in both symptomatic and asymptomatic patients, thereby enhancing clinical screening.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The Institutional Review Board (IRB) of China Medical University Hospital approved this retrospective, single-center study (IRB number CMUH110-REC3-019). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because The present study was under a retrospective, single-center study, waived by the institutional review board (IRB) of China Medical University Hospital and approved by approval number CMUH110-REC3-019.

Author contributions

H-YL: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. K-CH: Supervision, Writing – original draft. S-YC: Data curation, Formal analysis, Supervision, Writing – original draft. C-YY: Data curation, Formal analysis, Supervision, Writing – original draft. T-HS: Data curation, Formal analysis, Supervision, Writing – original draft. M-HL: Data curation, Formal analysis, Supervision, Writing – original draft. KN: Formal analysis, Writing – original draft.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. The study was partly funded by China Medical University Hospital, China Medical University, Taiwan (DMR-104-010).

Acknowledgments

Enago Academic editing edited this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2025.1496109/full#supplementary-material>

References

- Adedinsowo, D., Carter, R. E., Attia, Z., Johnson, P., Kashou, A. H., Dugan, J. L., et al. (2020). Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circ. Arrhythm. Electrophysiol.* 13:e008437. doi: 10.1161/CIRCEP.120.008437
- Ahmed, H. M., Al-Mallah, M. H., McEvoy, J. W., Nasir, K., Blumenthal, R. S., Jones, S. R., et al. (2015). Maximal exercise testing variables and 10-year survival: fitness risk score derivation from the FIT project. *Mayo Clin. Proc.* 90, 346–355. doi: 10.1016/j.mayocp.2014.12.013
- Alzubi, J. A., Alzubi, O. A., Beseiso, M., Budati, A. K., and Shankar, K. (2021). Optimal multiple key-based homomorphic encryption with deep neural networks to secure medical data transmission and diagnosis. *Expert. Syst.* 39:e12879. doi: 10.1111/essy.12879
- Angeja, B. G., de Lemos, J., Murphy, S. A., Marble, S. J., Antman, E. M., Cannon, C. P., et al. (2002). Impact of diabetes mellitus on epicardial and microvascular flow after fibrinolytic therapy. *Am. Heart J.* 144, 649–656. doi: 10.1067/mhj.2002.124869
- Arbab-Zadeh, A., and Hoe, J. (2011). Quantification of coronary arterial stenoses by multidetector CT angiography in comparison with conventional angiography methods, caveats, and implications. *JACC Cardiovasc. Imaging* 4, 191–202. doi: 10.1016/j.jcmg.2010.10.011
- Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., et al. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 394, 861–867. doi: 10.1016/S0140-6736(19)31721-0
- Babaoglu, I., Baykan, O. K., Aygul, N., Ozdemir, K., and Bayrak, M. (2009). Assessment of exercise stress testing with artificial neural network in determining coronary artery disease and predicting lesion localization. *Expert Syst. Appl.* 36, 2562–2566. doi: 10.1016/j.eswa.2007.11.013
- Babaoglu, I., Findik, O., and Bayrak, M. (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. *Expert Syst. Appl.* 37, 2182–2185. doi: 10.1016/j.eswa.2009.07.055
- Banerjee, R., Ghose, A., and Mandana, K. M. A Hybrid CNN-LSTM Architecture for Detection of Coronary Artery Disease from ECG. (2020) International joint conference on neural networks (IJCNN); 19–24 Jul 2020; Glasgow, UK: IEEE; 2020.
- Benkarim, O., Paquola, C., Park, B. Y., Kebets, V., Hong, S. J., Vos de Wael, R., et al. (2022). Population heterogeneity in clinical cohorts affects the predictive accuracy of brain imaging. *PLoS Biol.* 20:e3001627. doi: 10.1371/journal.pbio.3001627
- Bhatt, D. Cardiovascular Intervention: a companion to Braunwald's Heart Disease. (2015).
- Chen, K. W., Wang, Y. C., Liu, M. H., Tsai, B. Y., Wu, M. Y., Hsieh, P. H., et al. (2022). Artificial intelligence-assisted remote detection of ST-elevation myocardial infarction using a mini-12-lead electrocardiogram device in prehospital ambulance care. *Front. Cardiovasc. Med.* 9:1001982. doi: 10.3389/fcvm.2022.1001982
- Cheng, J., Zou, Q., and Zhao, Y. (2021). ECG signal classification based on deep CNN and BiLSTM. *BMC Med. Inform. Decis. Mak.* 21:365. doi: 10.1186/s12911-021-01736-y
- Christman, M. P., Bittencourt, M. S., Hulten, E., Saksena, E., Hainer, J., Skali, H., et al. (2014). Yield of downstream tests after exercise treadmill testing: a prospective cohort study. *J. Am. Coll. Cardiol.* 63, 1264–1274. doi: 10.1016/j.jacc.2013.11.052
- Detrano, R., Gianrossi, R., and Froelicher, V. (1989). The diagnostic accuracy of the exercise electrocardiogram: a meta-analysis of 22 years of research. *Prog. Cardiovasc. Dis.* 32, 173–206. doi: 10.1016/0033-0620(89)90025-x
- Fletcher, G. F., Ades, P. A., Kligfield, P., Arena, R., Balady, G. J., Bittner, V. A., et al. (2013). Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation* 128, 873–934. doi: 10.1161/CIR.0b013e31829b5b44
- Gencbay, M., Degertekin, M., Ermeidan, C., Unalp, A., and Turan, F. (1999). Exercise electrocardiography test in patients with aortic stenosis. Differential features from that of coronary artery disease. *Int. J. Cardiol.* 69, 281–287. doi: 10.1016/s0167-5273(99)00054-6
- Ghaffari, S., Asadzadeh, R., Tajlil, A., Mohammadalian, A., and Pourafkari, L. (2017). Predictive value of exercise stress test-induced ST-segment changes in leads V(1) and avR in determining angiographic coronary involvement. *Ann. Noninvasive Electrocardiol.* 22:e12370. doi: 10.1111/anec.12370
- Gibbons, R. J., Balady, G. J., Bricker, J. T., Chaitman, B. R., Fletcher, G. F., Froelicher, V. F., et al. (2002). ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association task force on practice guidelines (committee to update the 1997 exercise testing guidelines). *J. Am. Coll. Cardiol.* 40, 1531–1540. doi: 10.1016/s0735-1097(02)02164-2
- Huang, Z. A., Sang, Y., Sun, Y., and Lv, J. (2023). Neural network with a preference sampling paradigm for imbalanced data classification. *IEEE Trans. Neural. Netw. Learn. Syst.* PP. doi: 10.1109/TNNLS.2022.3231917
- Huang, P. S., Tseng, Y. H., Tsai, C. F., Chen, J. J., Yang, S. C., Chiu, F. C., et al. (2022). An artificial intelligence-enabled ECG algorithm for the prediction and localization of angiography-proven coronary artery disease. *Biomedicine* 10:394. doi: 10.3390/biomedicine10020394
- Kiranyaz, S., Ince, T., Hamila, R., and Gabbouj, M. (2015). Convolutional neural networks for patient-specific ECG classification. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2015, 2608–2611. doi: 10.1109/EMBC.2015.7318926
- Knuuti, J., Wijns, W., Saraste, A., Capodanno, D., Barbato, E., Funck-Brentano, C., et al. (2020). 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur. Heart J.* 41, 407–477. doi: 10.1093/eurheartj/ehz425
- Kose, U., Deperlioglu, O., Alzubi, J., and Patrut, B. (2021). Deep learning for medical decision support systems. *2020 Edition* Edn: Springer.
- Lee, Y. H., Tsai, T. H., Chen, J. H., Huang, C. J., Chiang, C. E., Chen, C. H., et al. (2022). Machine learning of treadmill exercise test to improve selection for testing for coronary artery disease. *Atherosclerosis* 340, 23–27. doi: 10.1016/j.atherosclerosis.2021.11.028
- Lehtinen, R. (1999). ST/HR hysteresis: exercise and recovery phase ST depression/heart rate analysis of the exercise ECG. *J. Electrocardiol.* 32, 198–204. doi: 10.1016/s0022-0736(99)90080-8
- Malakar, A. K., Choudhury, D., Halder, B., Paul, P., Uddin, A., and Chakraborty, S. (2019). A review on coronary artery disease, its risk factors, and therapeutics. *J. Cell. Physiol.* 234, 16812–16823. doi: 10.1002/jcp.28350
- Marwick, T. H., Torelli, J., Harjai, K., Haluska, B., Pashkow, F. J., Stewart, W. J., et al. (1995). Influence of left ventricular hypertrophy on detection of coronary artery disease using exercise echocardiography. *J. Am. Coll. Cardiol.* 26, 1180–1186. doi: 10.1016/0735-1097(96)81472-0
- Marzlin, K. M., and Webner, C. (2019). Chronotropic incompetence. *AACN Adv. Crit. Care* 30, 294–300. doi: 10.4037/aacnacc2019182
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature* 577, 89–94. doi: 10.1038/s41586-019-1799-6
- Mieres, J. H., Gulati, M., Bairey Merz, N., Berman, D. S., Gerber, T. C., Hayes, S. N., et al. (2013). Role of noninvasive testing in the clinical evaluation of women with suspected ischemic heart disease: a consensus statement from the American Heart Association. *Circulation* 130, 350–379. doi: 10.1161/CIR.0000000000000061
- Movassagh, A. A., Alzubi, J. A., Gheisari, M., Rahimi, M., Mohan, S., Abbasi, A. A., et al. (2013). Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model. *J. Ambient. Intell. Humaniz. Comput.* 14, 6017–6025. doi: 10.1007/s12652-020-02623-6
- Nallamothu, B. K., Spertus, J. A., Lansky, A. J., Cohen, D. J., Jones, P. G., Kureshi, F., et al. (2013). Comparison of clinical interpretation with visual assessment and quantitative coronary angiography in patients undergoing percutaneous coronary

intervention in contemporary practice: the assessing angiography (A2) project. *Circulation* 127, 1793–1800. doi: 10.1161/CIRCULATIONAHA.113.001952

Okin, P. M., Devereux, R. B., Lee, E. T., Galloway, J. M., Howard, B. V., and Strong, H. S. (2004). Electrocardiographic repolarization complexity and abnormality predict all-cause and cardiovascular mortality in diabetes: the strong heart study. *Diabetes* 53, 434–440. doi: 10.2337/diabetes.53.2.434

O'Neal, W. T., Lee, K. E., Soliman, E. Z., Klein, R., and Klein, B. E. (2017). Predictors of electrocardiographic abnormalities in type 1 diabetes: the Wisconsin epidemiologic study of diabetic retinopathy. *J. Endocrinol. Investig.* 40, 313–318. doi: 10.1007/s40618-016-0564-z

Sato, Y., Kawakami, R., Sakamoto, A., Cornelissen, A., Mori, M., Kawai, K., et al. (2022). Sex differences in coronary atherosclerosis. *Curr. Atheroscler. Rep.* 24, 23–32. doi: 10.1007/s11883-022-00980-5

Schultz, M. G., La Gerche, A., and Sharman, J. E. (2017). Blood pressure response to exercise and cardiovascular disease. *Curr. Hypertens. Rep.* 19:89. doi: 10.1007/s11906-017-0787-1

Shaw, L. J., Peterson, E. D., Shaw, L. K., Kesler, K. L., DeLong, E. R., Harrell, F. E., et al. (1998). Use of a prognostic treadmill score in identifying diagnostic coronary disease subgroups. *Circulation* 98, 1622–1630. doi: 10.1161/01.CIR.98.16.1622

Siegler, J. C., Rehman, S., Bhumireddy, G. P., Abdula, R., Klem, I., Brener, S. J., et al. (2011). The accuracy of the electrocardiogram during exercise stress test based on heart size. *PLoS One* 6:e23044. doi: 10.1371/journal.pone.0023044

Snader, C. E., Marwick, T. H., Pashkow, F. J., Harvey, S. A., Thomas, J. D., and Lauer, M. S. (1997). Importance of estimated functional capacity as a predictor of all-cause mortality among patients referred for exercise thallium single-photon emission computed tomography: report of 3,400 patients from a single center. *J. Am. Coll. Cardiol.* 30, 641–648. doi: 10.1016/s0735-1097(97)00217-9

Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., and Roth, G. A. (2022). The global burden of cardiovascular diseases and risk: a compass for future health. Washington DC: American College of Cardiology Foundation, 2361–2371.

Verma, D. and Agarwal, S. Cardiac arrhythmia detection from single-lead ECG using CNN and LSTM assisted by oversampling. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 19–22 September 2018; Bangalore, India: IEEE; (2018).

Vrints, C., Andreotti, F., Koskinas, K. C., Rossello, X., Adamo, M., Ainslie, J., et al. (2024). 2024 ESC guidelines for the management of chronic coronary syndromes. *Eur. Heart J.* 45, 3415–3537. doi: 10.1093/eurheartj/ehae177

Yadav, S. S., and Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* 6:113. doi: 10.1186/s40537-019-0276-2

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. Hierarchical attention networks for document classification. 2016 conference of the North American chapter of the association for computational linguistics: human language technologies; (2016).

Yildirim, O., Plawiak, P., Tan, R. S., and Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* 102, 411–420. doi: 10.1016/j.compbiomed.2018.09.009

Yilmaz, A., Hayiroglu, M. I., Salturk, S., Pay, L., Demircali, A. A., Coskun, C., et al. (2023). Machine learning approach on high risk treadmill exercise test to predict obstructive coronary artery disease by using P, QRS, and T waves' features. *Curr. Probl. Cardiol.* 48:101482. doi: 10.1016/j.cpcardiol.2022.101482

Zhang, J., Liu, A., Gao, M., Chen, X., Zhang, X., and Chen, X. (2020). ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. *Artif. Intell. Med.* 106:101856. doi: 10.1016/j.artmed.2020.101856

Zihlmann, M., Perekestenko, D., and Tschannen, M. (2017). Convolutional recurrent neural networks for electrocardiogram classification. 2017 computing in cardiology (CinC); 24–27 September 2017; Rennes, France: IEEE.



OPEN ACCESS

EDITED BY

Rui Providencia,
University College London, United Kingdom

REVIEWED BY

Habib Khan,
Western University, Canada
Sung Il Im,
Kosin University, Republic of Korea

*CORRESPONDENCE

Hao Liang
✉ lianghao@hnu.cm.edu.cn
Wen Zhang
✉ 46890358@qq.com

RECEIVED 22 February 2025

ACCEPTED 02 June 2025

PUBLISHED 19 June 2025

CITATION

Yin Z, Liu C, Xie C, Nie Z, Wei J, Zhang W and
Liang H (2025) Identification of atrial fibrillation
using heart rate variability: a meta-analysis.
Front. Cardiovasc. Med. 12:1581683.
doi: 10.3389/fcvm.2025.1581683

COPYRIGHT

© 2025 Yin, Liu, Xie, Nie, Wei, Zhang and Liang.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Identification of atrial fibrillation using heart rate variability: a meta-analysis

Ziwei Yin¹, Changxin Liu¹, Chenggong Xie², Zixing Nie³,
Jiaming Wei¹, Wen Zhang^{4*} and Hao Liang^{1*}

¹Institute of TCM Diagnostics, Hunan University of Chinese Medicine, Changsha, Hunan, China, ²School of Acupuncture and Tui-na and Rehabilitation, Hunan University of Chinese Medicine, Changsha, Hunan, China, ³Cardiology Department, The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, China, ⁴Geriatrics Department, The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, China

Background: Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and is associated with significant cardiovascular complications. Recently, artificial intelligence (AI) algorithms have leveraged heart rate variability (HRV) patterns to enhance the accuracy of AF identification.

Methods: We conducted a systematic review of the literature by searching four major biomedical databases—PubMed, Web of Science, Embase, and Cochrane Library—spanning from their inception to December 13, 2024, following the PRISMA guidelines. We extracted data on true positives, false positives, true negatives, and false negatives from the included studies, which were then synthesized to evaluate sensitivity and specificity comprehensively.

Results: Our final analysis included 12 diagnostic studies. Hierarchical summary receiver operating characteristic modeling revealed excellent discriminative ability, with a pooled sensitivity of 0.94 and specificity of 0.97. In detecting AF, the AI model demonstrated exceptional performance (sensitivity = 0.96, specificity = 0.99, AUC = 1.00). Subgroup analyses revealed that both deep learning algorithms (sensitivity = 0.95, specificity = 0.98, AUC = 0.99) and multi-database studies (sensitivity = 0.96, specificity = 0.97, AUC = 0.99) demonstrated enhanced accuracy in AF identification compared to other approaches.

Conclusion: Machine learning can effectively identify AF with HRV in ECG, especially in diagnosis and detection, with deep learning algorithms and multiple-databases outperforming other diagnostic methods.

Systematic Review Registration: <https://www.crd.york.ac.uk/PROSPERO/>, PROSPERO (CRD42025634406).

KEYWORDS

deep learning, machine learning, atrial fibrillation, heart rate variability, ECG, meta-analysis

1 Introduction

Atrial fibrillation (AF) is the most common persistent arrhythmia encountered in clinical practice, and its global disease burden continues to increase as the acceleration of aging population (1). Data from the Framingham Heart Study revealed a threefold increase in the incidence of AF over the past 50 years, underscoring its growing impact as a significant public health concern, particularly for the elderly (2). Currently, the clinical diagnosis of AF largely depends on patient-reported symptoms and electrocardiogram (ECG) results. However, existing evidence suggests that these traditional methods are associated with a misdiagnosis rate of approximately 20% (3). This diagnostic uncertainty may result in

excessive treatment, increasing patient burdens and inefficient resource allocation. As a result, the development of innovative and accurate diagnostic technologies has become a critical priority to enhance the diagnostic approach for AF.

Heart rate variability (HRV), which quantifies fluctuations in beat-to-beat intervals, has become an established tool in clinical prediction models for sudden cardiac death and life-threatening arrhythmia (4). Notably, recent advancements integrating wearable or implantable HRV data with artificial intelligence (AI)-based analytical systems have facilitated novel strategies for the early detection of AF and precision risk stratification (5, 6).

AI systems replicate human cognitive processes through autonomous decision-making architectures, with their inherent strength rooted in hierarchical pattern recognition and deep processing of complex datasets. In the field of cardiovascular medicine, machine learning (ML)-enhanced ECG interpretation has demonstrated measurable improvements in predictive performance: Alimbayeva et al. established a cardiovascular risk stratification model through multimodal integration of ECG biomarkers using logistic regression, random forest classifiers, and convolutional neural networks (7). While, Khurshid's group developed an ML-driven framework synergizing ECG patterns with clinical risk factors, achieving significant predictive capacity for the onset of AF (8). These developments, driven by ongoing technological advancements and the growing availability of open-access clinical data, positions AI-driven approaches as powerful tools to enhance the accuracy of AF detection and improve diagnostic efficiency.

The increasing clinical adoption of HRV monitoring technologies has driven substantial research interest in AI-driven HRV feature engineering for AF prediction. However, existing studies demonstrate substantial methodological heterogeneity in algorithm architectures, data quality standards, and validation protocols, which may introduce potential biases in diagnostic performance evaluations. This study aims to fill this gap by conducting the first diagnostic test accuracy meta-analysis that simultaneously evaluates both sensitivity and specificity of ML-enhanced HRV analysis for AF detection. The findings provide essential insights to inform clinical decision-making and offer valuable guidance for future algorithmic improvements through standardized performance benchmarking.

2 Materials and methods

2.1 Protocol and registration

This meta-analysis is reported according to the Preferred Reporting Items for Meta-Analyses (PRISMA) statement (Supplementary Table S1) (9), and it was registered in the PROSPERO database (CRD42025634406).

2.2 Search strategy and study selection

A systematic search was conducted across four major databases including PubMed, Web of Science, Embase, and the Cochrane Library, from their inception through December 13, 2024, limited to English-language publications. The search strategy is as follows:

(atrial fibrillation OR auricular fibrillation) and (heart rate variability OR HRV OR SDNN OR SDANN OR RMSSD) and (artificial intelligence OR machine learning OR deep learning).

Two researchers (YZW and LCX) completed the literature screening separately—title and abstract screening eliminated clearly irrelevant records (e.g., non-AI methods, non-ECG data, or animal studies)—followed by full-text evaluation of potentially eligible articles against predefined criteria. Discrepancies were resolved through consensus discussions or third-reviewer arbitration (LH).

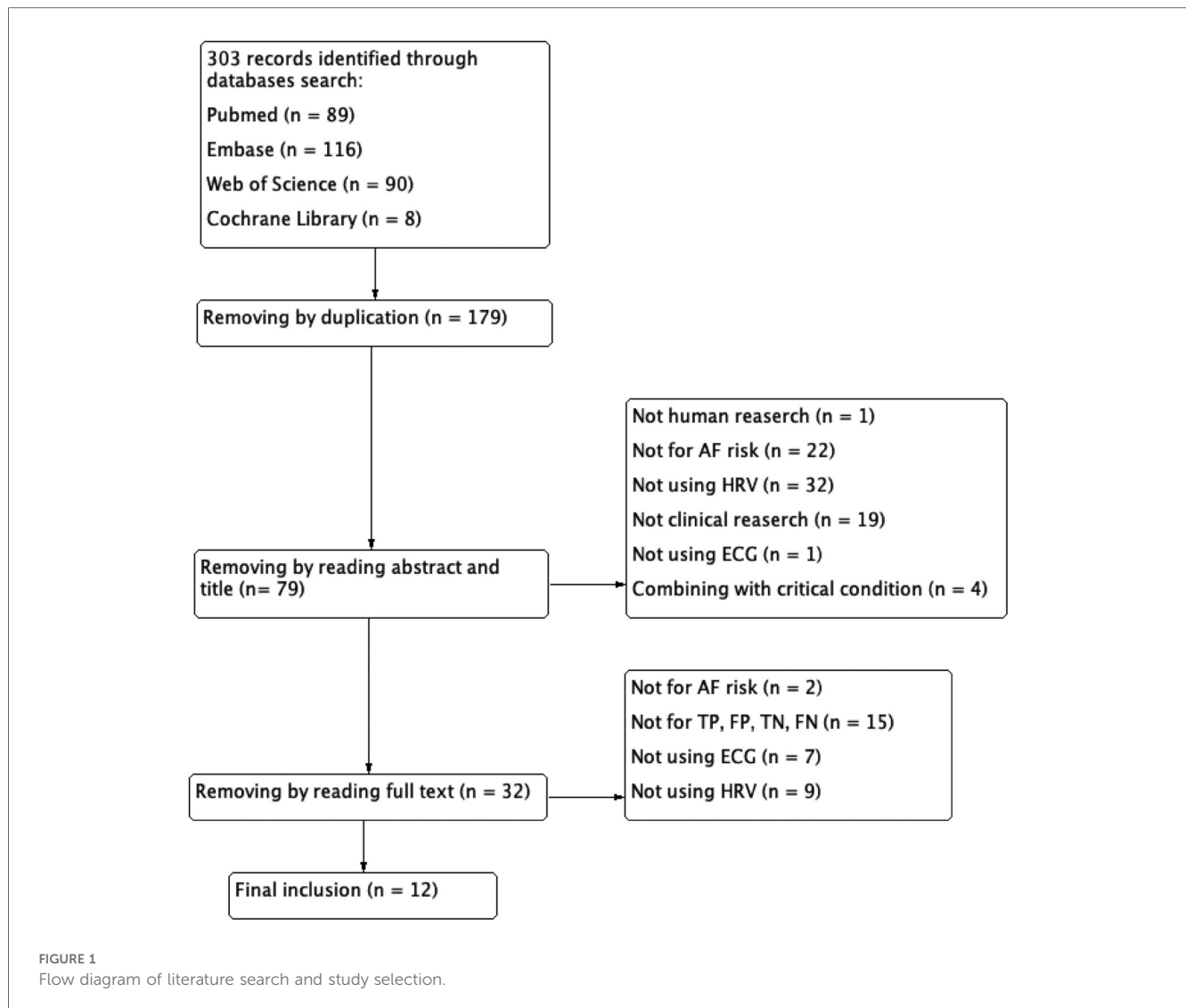
Inclusion criteria were as follows—(1) implementation of machine learning algorithms, (2) using ECG-derived data, (3) AF as the primary clinical endpoint, (4) human clinical studies, and (5) the prediction of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) either be included in studies or can be calculated; Exclusion criteria included—(1) duplication publications, (2) studies involving critically ill populations, and (3) undefined AI methodologies. The complete selection process was detailed in Figure 1.

2.3 Data extraction and quality assessment

Two investigators (YZW and LCX) independently extracted data using standardized templates. The following parameters were collected: first author, publication year, data source, specifications of the AI algorithm, and diagnostic performance metrics (sensitivity, specificity). The methodological quality was assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (10), which evaluates four critical domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. After the initial data extraction, cross-validation was conducted to verify inter-rater consistency. Due to the limited sample size ($n < 20$), Cohen's kappa coefficients were not calculated, as agreement measures can be statistically unstable with small sample sizes.

2.4 Statistical analysis

When a patient was diagnosed as AF, it was considered a positive result; conversely, when diagnosed as non-AF, it was considered a negative result. So, TP refers to the number of true positive samples, which are correctly predicted as positive. FP indicates the number of false positive samples, which are negative but incorrectly predicted as positive. TN refers to the number of true negative samples, which are correctly predicted as negative. FN represents the number of false negative samples, which are positive but incorrectly predicted as negative. Two independent investigators (YZ and LC) extracted the number of TP, FP, FN, and TN from each eligible study. We implemented bivariate random-effects modeling coupled with hierarchical summary receiver operating characteristic analysis to derive pooled sensitivity and specificity estimates with 95% confidence intervals (CI). Heterogeneity assessment incorporated Cochran's Q and I^2 statistics, with predefined thresholds: Q-test P -value ≥ 0.10 and $I^2 \leq 50\%$ indicating low heterogeneity, vs. $P < 0.10$ and $I^2 > 50\%$



denoting substantial heterogeneity. For analyses demonstrating substantial heterogeneity, we conducted meta-regression, sensitivity analyses, leave-one-out analyses. And we conducted subgroup analysis with the type of AI algorithm and number of adopted data sets as parameters to investigate potential sources of variability.

Publication bias was evaluated using Deek's funnel plot asymmetry test, with statistical significance set at $P > 0.05$ indicating absence of bias. Finally, we conducted a clinical diagnostic test for this diagnostic strategy, calculating both the positive likelihood ratio (PLR) and the negative likelihood ratio (NLR) to assess its diagnostic accuracy. All statistical computations were performed in Stata/MP 18.0 and R 4.4.2.

3 Results

3.1 Characteristics of included studies

This analysis incorporated twelve diagnostic studies (11, 22), five of these studies predicted the occurrence of AF (11, 14, 16,

18, 19), and another seven used AI algorithms to detect AF (12, 13, 15, 17, 20–22). And the included studies with algorithm type distributed as follows: four studies employed DL methods (11, 12, 15, 17), seven utilized ML approaches (13, 14, 18–22), and only one study incorporated both two algorithmic ways (16). Data mainly originated from the PhysioNet platform (<https://physionet.org/>), MIT-BIH Atrial Fibrillation Database, and MIT-BIH Arrhythmia Database. Only one investigation utilized clinical datasets (18), while another synthesized clinical and repository data (12). Regarding data diversity, only five original studies used multi-source datasets (11–13, 16, 18), other 7 studies relying on single-source one. Comprehensive baseline characteristics were presented in Table 1.

3.2 Quality assessment

The risk of bias in the included studies was evaluated using the QUADAS-2 tool. The evaluation results are as follows: (1) patient election (2 studies with high risk of bias, 6 with unclear risk of bias,

TABLE 1 Main characteristics of the 12 included researches in this meta-analysis.

Authors	Database	HRV parameter	Methodology	Performance
Chen et al. (11)	Atrial Fibrillation Paroxysmal Database MIT-BIH Atrial Fibrillation Database MIT-BIH Normal Sinus Rhythm Database	R-R interval	Convolutional Neural Network	SEN = 0.9712 SPE = 0.9777
Tutuko et al. (12)	MIT-BIH Atrial Fibrillation Database 2017 PhysioNet/CinC Challenge Database 2018 PhysioNet/CinC Challenge Database ECG Long Term AF Database Atrial Fibrillation Paroxysmal Database MIT-BIH Arrhythmia Database AF Termination Challenge Database Fantasia Database ECG recording from Chapman University and Shaoxing People's Hospital ECG recording from an Indonesian Hospital	R-R interval	Convolutional Neural Network	SEN = 0.9980 SPE = 0.9980
Udawat and Singh (13)	MIT-BIH Atrial Fibrillation Database MIT-BIH Arrhythmia Database	R-R interval	Fourier Decomposition Method	SEN = 0.9940 SPE = 0.9950
Wu et al. (14)	Atrial Fibrillation Paroxysmal Database	^a 11 time domain parameter 7 frequency domain parameter 7 nonlinear parameter	Bagging Ensemble Learning Method AdaBoost Ensemble Learning Method Stacking Ensemble Learning Method	SEN = 0.8800 SPE = 0.9600
Marinucci et al. (15)	2017 PhysioNet/CinC Challenge Database	MRR, SDRR, RMSRR, PRR50	Artificial Neural Network	SEN = 0.8120 SPE = 0.8120
Chesnokov (16)	Atrial Fibrillation Paroxysmal Database MIT-BIH Atrial Fibrillation Database	pVLF, pLF, pHF, LF/HF, SampEn, ApEn, MSE, and MAE	Artificial Neural Network	SEN = 0.6818 SPE = 1.0000
			Support Vector Machine (Radial Basis Function Kernel)	SEN = 0.8372 SPE = 0.7647
			Support Vector Machine (Sigmoid Kernel)	SEN = 0.8372 SPE = 0.7647
Sanjana et al. (17)	2017 PhysioNet/CinC Challenge Database	MRR, SDNN, RMSSD	Recurrent Neural Network	SEN = 0.9034 SPE = 0.9687
			Gated Recurrent Unit	SEN = 0.8725 SPE = 0.9787
Saiz-Vivo et al. (18)	Reveal LINQ usability study (NCT01965899) Single Center Clinical Trail (29)	MRR, pNN50, pNN20, RMSSD, SDNN, TINN, TRI, ApEn, SampEn, SD1, SD2, SD1/SD2, DFAa1a2	Support Vector Machine	SEN = 0.8275 SPE = 0.5950
Xin and Zhao (19)	Atrial Fibrillation Paroxysmal Database	^b 4 time domain parameter 4 frequency domain parameter	Multi-scale Wavelet α -entropy	SEN = 0.9488 SPE = 0.8948
Asl et al. (20)	MIT-BIH Arrhythmia Database	R-R interval	Generalized Discriminant Analysis Support Vector Machine	SEN = 0.9463 SPE = 0.9972
Mei et al. (21)	2017 PhysioNet/CinC Challenge Database	R-R interval	Support Vector Machine Bagging Trees	SEN = 0.8840 SPE = 0.9958
Bus et al. (22)	Long-Term Atrial Fibrillation Database	pRRx	Fourier Decomposition Method	SEN = 0.9042 SPE = 0.9537

^aTime domain parameter: MRR, SDNN, HR, SDHR, minHR, maxHR, RMSSD, NN50, pNN50, HRV triangular index, and TINN, Frequency domain parameter: pVLF, pLF, pHF, LF/HF, total spectral power, LF/(TP-VLF), and HF/(TP-VLF), Nonlinear parameter: SD1, SD2, SD2/SD1, ApEn, SampEn, and short-term and long-term fluctuations of DFA.

^bTime domain parameter: MRR, SDNN, RMSSD, and pNN50, Frequency domain parameter: pVLF, pLF, pHF, and LF/HF.

MRR: mean of RR interval, SDNN: standard deviation of normal to normal RR intervals, HR: heart rate, SDHR: standard deviation of instantaneous heart rate values, minHR: min heart rate per minute, maxHR: maximum heart rate per minute, RMSSD: root mean square of successive RR interval differences, NN50: number of successive RR interval pairs that differ more than 50 ms, pNN50: NN50 divided by the total number of all NN intervals, pNN20: NN20 divided by the total number of all NN intervals, TINN: baseline width of the NN interval histogram, TRI: triangular index, pVLF: absolute power of VLF band, pLF: absolute power of LF band, pHF: absolute power of HF band, LF/HF: ratio between LF and HF band powers, LF/(TP-VLF): normalized LFP, HF/(TP-VLF): normalized HFP, SD1: poincaré plot standard deviation perpendicular the line of identity, SD2: poincaré plot standard deviation along the line of identity, SD2/SD1: ratio of SD2 to SD1, ApEn: approximate entropy, SampEn: sample entropy, DFAa1a2: short-term and long-term fluctuations of detrended fluctuation analysis, MSE: multiscale sample entropy, MAE: multiscale approximate entropy, pRRx: percentage of successive RR intervals differing by at least x ms, AF: atrial fibrillation, ECG: electrocardiogram; SEN: sensitivity, SPE: specificity.

and 4 with low risk of bias), (2) index test (all 12 studies with low risk of bias), (3) reference standard (no study with high risk of bias, 2 with unclear risk of bias, 10 with low risk of bias), and (4) flow and timing (4 studies with high risk of bias, 5 with unclear risk of bias, and 3 with low risk of bias). A detailed summary of the quality assessment results can be found in Figure 2.

3.3 Meta analysis

A total of 12 studies were included in the meta-analysis. The forest plot indicated that AI exhibited high diagnostic performance in identifying AF with HRV from ECG. The pooled sensitivity was 0.94 (95% CI: 0.87–0.98), while the pooled

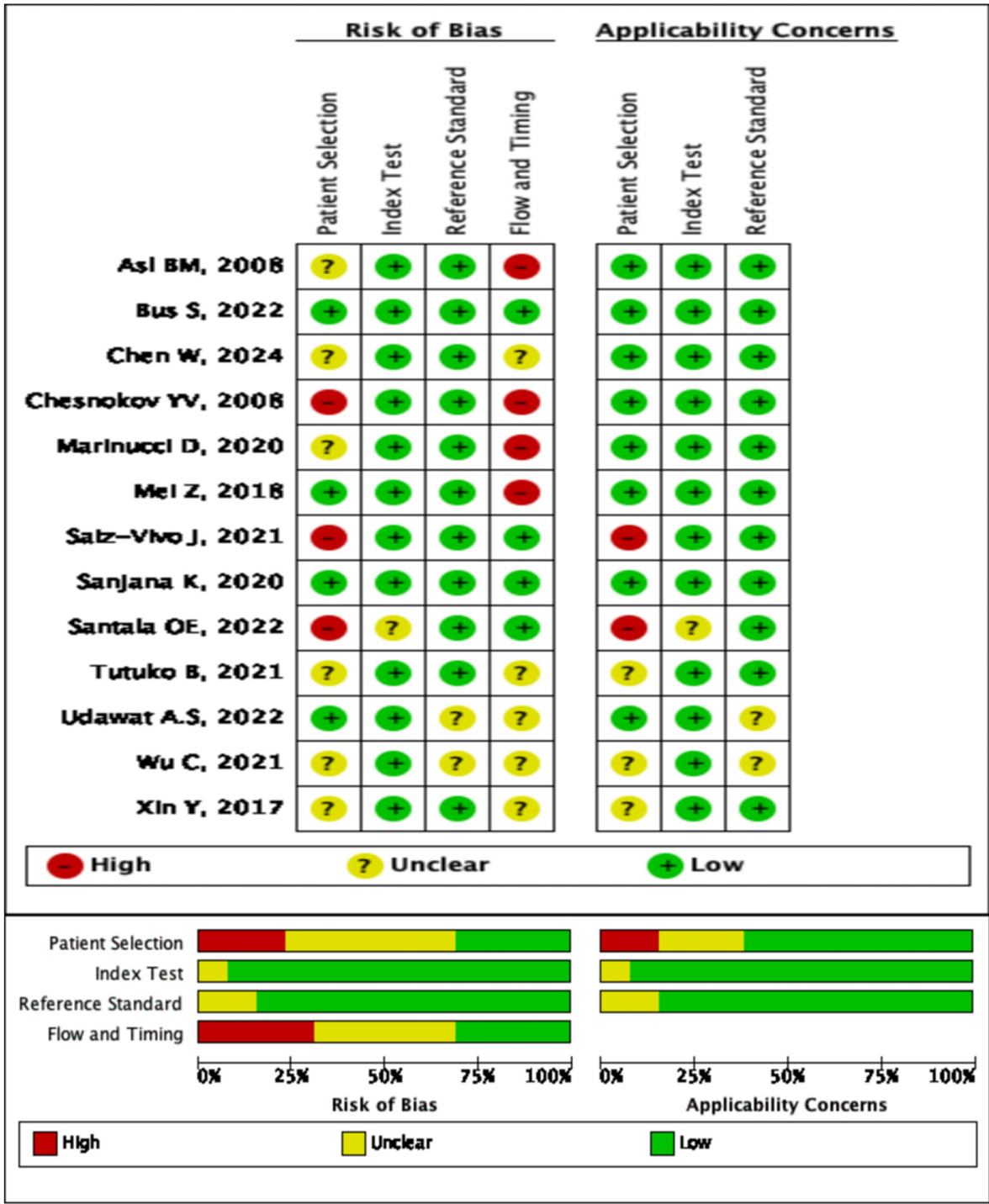


FIGURE 2
Summary of risk of bias across all included studies.

specificity was 0.97 (95% CI: 0.92–0.99). The summary receiver operating characteristic curve showed an area under the curve (AUC) of 0.99 (95% CI: 0.97–0.99), with most studies demonstrating strong sensitivity and specificity (Figure 3A). In predicting the onset of AF, the AI algorithm demonstrated a sensitivity of 0.87 (95% CI: 0.74–0.94), specificity of 0.90 (95% CI: 0.72–0.97), and an AUC of 0.94 (Figure 3B). For AF

detection, the AI algorithm exhibited even superior performance, with sensitivity of 0.96 (0.87–0.99), specificity of 0.99 (0.96–1.00), and AUC of 1.00 (0.99–1.00) (Figure 3C). The above three results are summarized in Table 2. However, considerable heterogeneity was observed in the forest plot, with sensitivity showing an I^2 of 99.95%, $Q=24,113.88$, and $P<0.1$, and specificity exhibiting an I^2 of 99.94%, $Q=18,584.74$, and

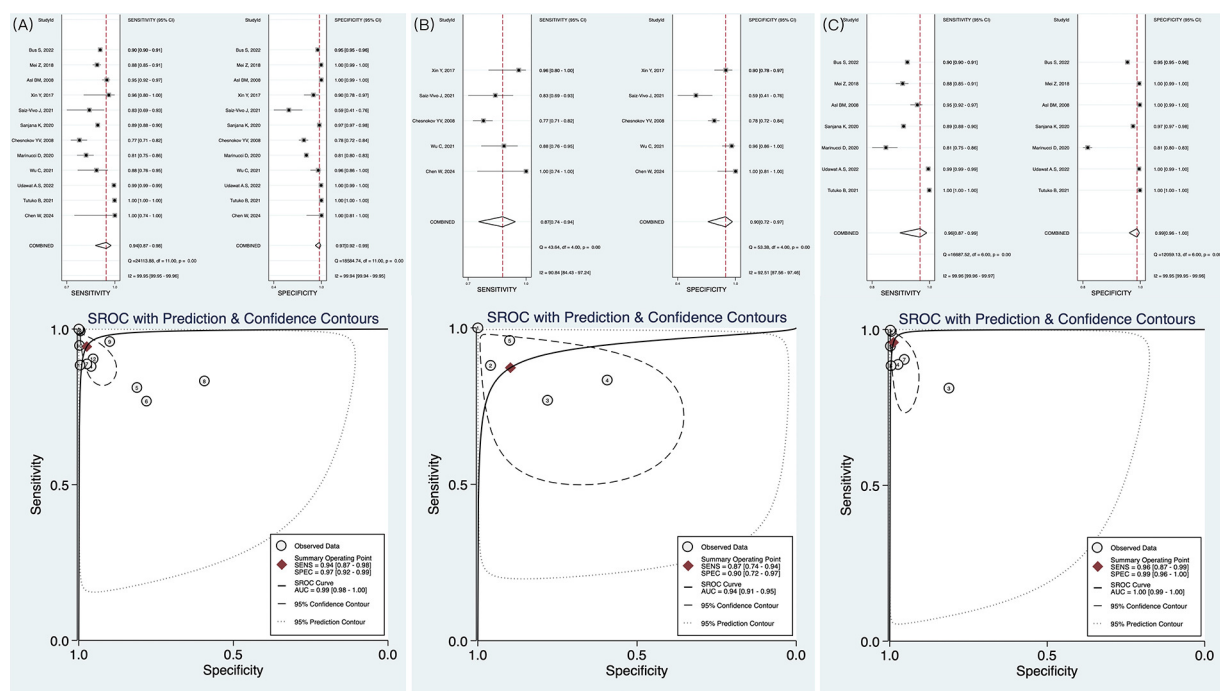


FIGURE 3 Forest plot and receiver operating characteristics curves of artificial intelligence for AF identification in HRV. (A) The result of pooled research; (B) The results of AI in predicting AF; (C) The results of AI in detecting AF.

TABLE 2 Main results of meta-analysis.

Research type	Sensitivity	Specificity	AUC
Pooled	0.94 (0.87–0.98)	0.97 (0.92–0.99)	0.99 (0.98–1.00)
Prediction of AF	0.87 (0.74–0.94)	0.90 (0.72–0.97)	0.94 (0.91–0.97)
Detection of AF	0.96 (0.87–0.99)	0.99 (0.96–1.00)	1.00 (0.99–1.00)

AF, atrial fibrillation; AUC, area under the curve.

$P < 0.1$. To explore the sources of heterogeneity, we conducted meta-regression, sensitivity analysis, and subgroup analysis.

3.4 Meta-regression and sensitivity analyses

Meta-regression analyses were performed based on the baseline characteristics of the included studies, focusing on two factors: AI algorithm type and the number of datasets used. The results indicated that the heterogeneity between different AI algorithm types was not statistically significant ($P > 0.05$), and no significant differences in specificity were observed between single-dataset and multi-dataset groups ($P > 0.05$). Therefore, neither the AI algorithm type nor the number of datasets explained the heterogeneity observed in the meta-analysis (Supplementary Figure S2). Sensitivity analysis demonstrated the robustness of the overall results (Supplementary Figure S3). When each study was individually excluded, neither the pooled effect size nor the heterogeneity exhibited any significant changes (Supplementary Table S4).

3.5 Subgroup analysis

Subgroup analysis revealed significant differences in diagnostic performance based on various AI algorithms type (Figure 4A). The DL model demonstrated nearly perfect discriminative ability, with an AUC of 0.99 (95% CI: 0.98–1.00), sensitivity of 0.95 (95% CI: 0.76–0.99), and specificity of 0.98 (95% CI: 0.93–1.00) (Supplementary Figure S5A). In contrast, AUC of ML models was 0.97 (95% CI: 0.96–0.98), with sensitivity of 0.92 (95% CI: 0.84–0.96) and specificity of 0.95 (95% CI: 0.84–0.99), which were slightly inferior to those of the DL model (Supplementary Figure S5B). Additionally, data diversity was found to have a crucial impact on model generalization ability (Figure 4B). When trained on a single database, AUC was 0.95 (95% CI: 0.92–0.96), with sensitivity of 0.89 (95% CI: 0.86–0.91) (Supplementary Figure S5C). However, cross-validation using multiple databases significantly improved diagnostic performance, with an AUC of 0.99 (95% CI: 0.98–1.00) and sensitivity increased to 0.96 (95% CI: 0.81–0.99) (Supplementary Figure S5D).

3.6 Clinical diagnostic testing

When the PLR value exceeds 10, a positive result significantly raises the likelihood of disease, shifting the pre-test probability from 20% to 90%. In this study, PLR was 37, suggesting this diagnostic marker effectively identified the presence of the

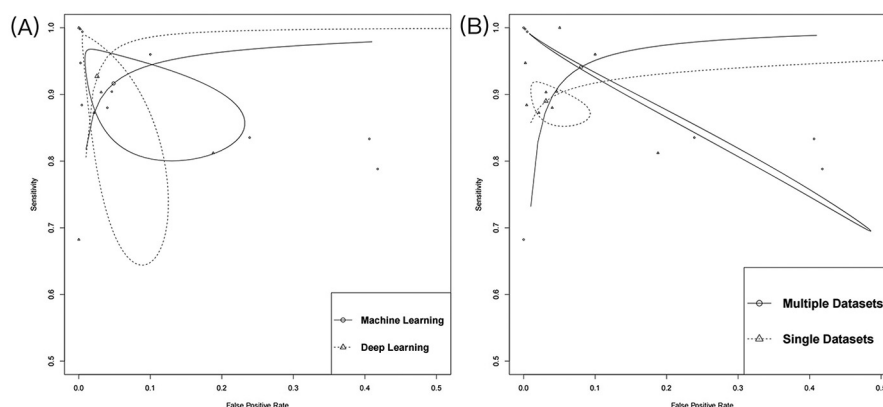


FIGURE 4

Summary of receiver operating characteristics. (A) The subgroup of algorithm type; (B) The subgroup of number of databases.

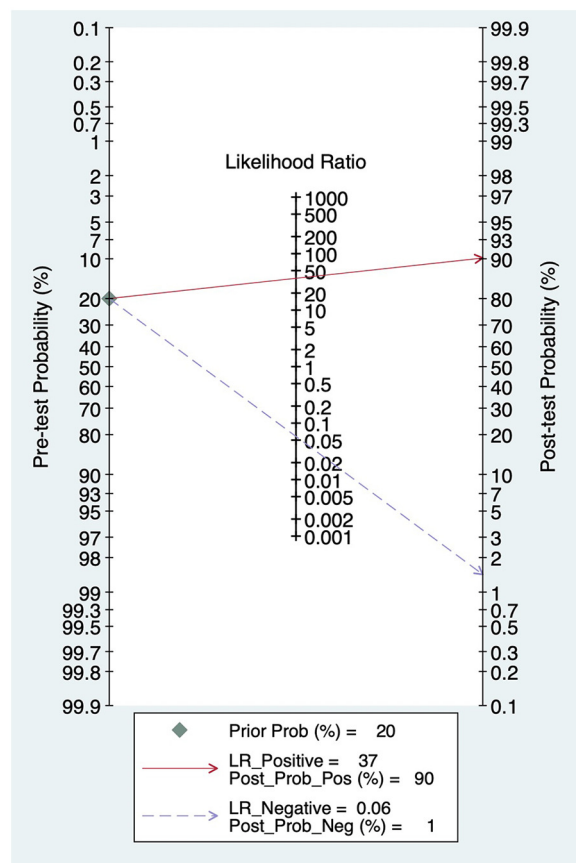


FIGURE 5

Fagan plot.

disease. Conversely, when the NLR is below 0.1, a negative result markedly reduces the probability of disease, lowering the pre-test probability from 20% to 1%. In this study, NLR was 0.06, demonstrating that this diagnostic marker is highly effective in ruling out the disease in negative cases (Figure 5).

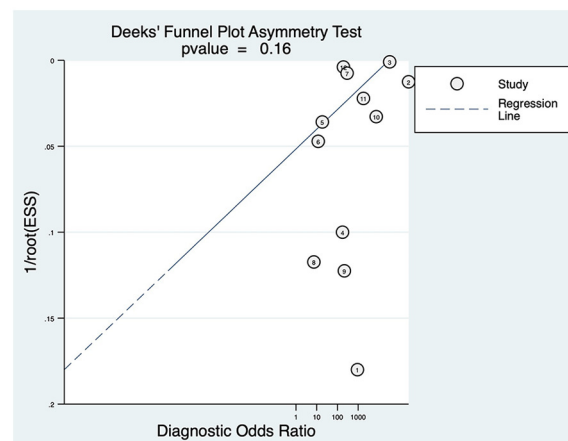


FIGURE 6

Deek's funnel plot.

3.7 Publication bias

Publication bias was evaluated using Deek's funnel plot asymmetry test. The results indicated a *P*-value of 0.16, suggesting that no publication bias was present in the studies included in this analysis (Figure 6).

4 Discussion

This study is the first to evaluate the diagnostic performance of HRV based on AI algorithms through a meta-analysis. The 12 included studies demonstrate that AI models exhibit outstanding performance in identifying AF, especially in detecting AF. The area under the summary receiver operating characteristic curve is the theoretical optimal value ($AUC = 1.00$), which seems to mean that AI could be the "gold method" for diagnosing AF. Notably, the clinical applicability analysis further validated the diagnostic

value of this technology: the PLR value is 37, and the NLR value is 0.06. These findings indicated that the AI-based HRV diagnostic tool has strong accuracy and rule out negative results effectively, highlighting its potential as a auxiliary tools for clinicians in identifying AF.

Our results are consistent with previous studies. For example, Manetas-Stavarakakis et al. conducted a cohort or case-control study using single-lead ECG to detect AF and reported a combined sensitivity of 92.3% and specificity of 96.2% (23). In comparison, our study found even higher sensitivity and specificity, suggesting that HRV may be a more accurate AF diagnostic marker. Subsequent subgroup analyses revealed that, compared to ML models, DL demonstrated superior diagnostic performance, which aligns with the findings of Solam Lee et al (24). Additionally, Xie C.G. et al. reached similar conclusions, with the DL model exhibiting significantly higher sensitivity (98.1%) compared to the ML model (91.5%) (25). DL models represent an important category of ML, encompassing algorithms such as deep feedforward neural networks, convolutional neural networks, and deep belief networks. A key feature of these models is their high degree of flexibility. Unlike ML models, the individual representations in DL are not manually designed by researchers but are instead learned from training data (26). Furthermore, DL models learn representations not in a single step, but across multiple layers between hidden layers. More importantly, the subsequent transformations between layers in DL models are non-linear, significantly enhancing the model's adaptability. This could explain the superior performance of the DL model observed in our study.

Although this study demonstrates promising combined results, several limitations are unavoidable. On the one hand, limitations stems from the included studies themselves. The majority of studies did not use external validation sets, and the widespread use of a single data source led to significant overlap between model training and validation sets. This overlap may hinder the model's ability to adapt to the complexity of real-world patients and clinical environments, potentially increasing misdiagnosis rates and medical risks. The AUC in detecting atrial fibrillation may imply an overlap between the training sets and test sets, and the lack of real-world external validation may be an important reason why the AUC is perfect. On the other hand, there was considerable heterogeneity observed across studies (sensitivity $I^2 = 99.95\%$, specificity $I^2 = 99.94\%$). Although meta-regression and sensitivity analyses confirmed the robustness of our findings, this suggested the presence of deeper underlying factors. One major cause of heterogeneity is the differences in the source and quality of the raw data. The majority of the included studies relied on the publicly available PhysioNet database, with only one study collecting clinical data and one combining both clinical and database data. While the standardized collection process of public databases reduces technical bias, it may not fully reflect the complexity of real-world clinical scenarios. Clinical data collection, is often susceptible to background noise (e.g., patients' movement). Unfortunately, due to the limited number of clinical diagnostic studies, we were unable to perform subgroup analysis to compare the diagnostic performance of

database vs. clinical data. Secondly, while all studies focused on HRV as the central feature, there were slight differences in the ECG features of HRV, leading to a lack of a standard procedures for HRV extraction. For example, Chen W. et al. used the RR interval as the HRV feature (11), while Bus S. et al. extracted the pRRx parameter to predict AF (22). Other studies incorporated multiple indicators as HRV features, including DNN, RMSSD, pNN50, and pNN20 (15, 18, 19). These technical variations led to inconsistent model inputs, affecting performance stability. Finally, since ECG signals are highly susceptible to various types of interference during data collection, background noise can obscure the true cardiac electrical activity, thereby affecting the performance and accuracy of AI models. Consequently, most diagnostic studies performed denoising prior to HRV signal input. Removing noise helps preserve crucial ECG signals (27), thereby reducing the rates of both false positives and false negatives. However, subtle AF fibrillation waves could be mistakenly classified as noise and eliminated, resulting in diagnostic bias. Additionally, there are significant differences in signal fidelity and computational efficiency among different denoising methods, which contributes to the substantial heterogeneity observed in this study.

The limitations of this study reflect common challenges in the current field. There is a clear need for standardized data collection processes and high-quality datasets to ensure consistency in data input. Additionally, it is crucial to identify and optimize the most effective HRV features to enhance the specificity of AF identification. More importantly, there is an urgent need for additional prospective studies to assess the real-world applicability of AI prediction models. Despite these challenges, the accuracy of AI models for identifying arrhythmias has already been shown to surpass that of general cardiologists (28). We look forward to the development of more comprehensive databases and more advanced AI algorithms, which can assist clinicians in better diagnosing atrial fibrillation.

5 Conclusion

In conclusion, AI effectively utilizes HRV in ECG signals to detect AF, with its DL algorithms and multi-database approaches demonstrating superior diagnostic performance.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding authors.

Author contributions

ZY: Writing – review & editing, Validation, Visualization, Writing – original draft. CL: Writing – original draft. CX: Writing – review & editing. ZN: Writing – review & editing. JW:

Writing – review & editing, Methodology. WZ: Writing – review & editing, Supervision. HL: Writing – review & editing, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the grants of Leading Research Program of Hunan University of Chinese Medicine (No.2022XJJ002), Science and Technology Innovation Program of Hunan Province (No. 2022RC1021), Key project of the Joint Fund of Hunan University of Chinese Medicine (2024XYLH339), General Project of Health Commission of Hunan Province, Project number (D202303019470).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Kornej J, Börschel C, Benjamin EJ, Schnabel RB. Epidemiology of atrial fibrillation in the 21st century, novel methods and new insights. *Circ Res.* (2020) 127(1):4–20. doi: 10.1161/CIRCRESAHA.120.316340
- Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. Fifty-year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the community. *Lancet.* (2015) 386(9989):154–62. doi: 10.1016/S0140-6736(14)61774-8
- Bogun F, Anh D, Kalahasty G, Wissner E, Serhal CB, Bazzi R, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med.* (2004) 117(9):636–42. doi: 10.1016/j.amjmed.2004.06.024
- Catai AM, Pastre CM, de Godoy MF, da Silva E, de Medeiros Takahashi AC, Vanderlei LCM. Heart rate variability: are you using it properly? Standardisation checklist of procedures. *Braz J Phys Ther.* (2020) 24(2):91–102. doi: 10.1016/j.bjpt.2019.02.006
- Oikawa J, Fukaya H, Ishii S, Nabeta T, Nakamura H, Ishizue N, et al. Variation in heart rate range by 24-h Holter monitoring predicts heart failure in patients with atrial fibrillation. *ESC Heart Fail.* (2022) 9(5):3092–100. doi: 10.1002/ehf2.14035
- Matusik PS, Matusik PT, Stein PK. Heart rate variability and heart rate patterns measured from wearable and implanted devices in screening for atrial fibrillation: potential clinical and population-wide applications. *Eur Heart J.* (2023) 44(13):1105–7. doi: 10.1093/eurheartj/ehac546
- Alimbayeva Z, Alimbayev C, Ozhikenov K, Bayanbay N, Ozhikenova A. Wearable ECG device and machine learning for heart monitoring. *Sensors.* (2024) 24(13):4201. doi: 10.3390/s24134201
- Alahdab F, Saad MB, Ahmed AI, Al Tashi Q, Aminu M, Han Y, et al. Development and validation of a machine learning model to predict myocardial blood flow and clinical outcomes from patients' electrocardiograms. *Cell Rep Med.* (2024) 5(10):101746. doi: 10.1016/j.xcrm.2024.101746
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br Med J.* (2009) 339:b2535. doi: 10.1136/bmj.b2535
- Whiting PF, Rutjes AWS, 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
- Chen W, Zheng P, Bu Y, Xu Y, Lai D. Achieving real-time prediction of paroxysmal atrial fibrillation onset by convolutional neural network and sliding window on R-R interval sequences. *Bioengineering (Basel).* (2024) 11(9):903. doi: 10.3390/bioengineering11090903
- Tutuko B, Nurmaini S, Tondas AE, Rachmatullah MN, Darmawahyuni A, Esafrir R, et al. AFibnet: an implementation of atrial fibrillation detection with convolutional neural network. *BMC Med Inform Decis Mak.* (2021) 21(1):216. doi: 10.1186/s12911-021-01571-1
- Udawat AS, Singh P. An automated detection of atrial fibrillation from single-lead ECG using HRV features and machine learning. *J Electrocardiol.* (2022) 75:70–81. doi: 10.1016/j.jelectrocard.2022.07.069
- Wu C, Hwang M, Huang TH, Chen YMJ, Chang YJ, Ho TH, et al. Application of artificial intelligence ensemble learning model in early prediction of atrial fibrillation. *BMC Bioinformatics.* (2021) 22(Suppl 5):93. doi: 10.1186/s12859-021-04000-2
- Marinucci D, Sbrollini A, Marcantoni I, Moretti M, Swenne CA, Burattini L. Artificial neural network for atrial fibrillation identification in portable devices. *Sensors (Basel).* (2020) 20(12):3570. doi: 10.3390/s20123570
- Chesnokov YV. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artif Intell Med.* (2008) 43(2):151–65. doi: 10.1016/j.artmed.2008.03.009
- Sanjana K, Sowmya V, Gopalakrishnan EA, Soman KP. Explainable artificial intelligence for heart rate variability in ECG signal. *Healthc Technol Lett.* (2020) 7(6):146–54. doi: 10.1049/htl.2020.0033
- Saiz-Vivo J, Corino VDA, Hatala R, de Melis M, Mainardi LT. Heart rate variability and clinical features as predictors of atrial fibrillation recurrence after catheter ablation: a pilot study. *Front Physiol.* (2021) 12:672896. doi: 10.3389/fphys.2021.672896
- Xin Y, Zhao Y. Paroxysmal atrial fibrillation recognition based on multi-scale wavelet α -entropy. *Biomed Eng Online.* (2017) 16(1):121. doi: 10.1186/s12938-017-0406-z
- Asl BM, Setarehdan SK, Mohebbi M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artif Intell Med.* (2008) 44(1):51–64. doi: 10.1016/j.artmed.2008.04.007
- Mei Z, Gu X, Chen H, Chen W. Automatic atrial fibrillation detection based on heart rate variability a spectral features. *IEEE ACCESS.* (2018) 6:53566–75. doi: 10.1109/ACCESS.2018.2871220
- Buś S, Jędrzejewski K, Guzik P. Statistical and diagnostic properties of pRRx parameters in atrial fibrillation detection. *J Clin Med.* (2022) 11(19):5702. doi: 10.3390/jcm11195702
- Manetas-Stavarakakis N, Sotiropoulou IM, Paraskevas T, Stavarakaki SM, Bampatsias D, Xanthopoulos A, et al. Accuracy of artificial intelligence-based

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2025.1581683/full#supplementary-material>

technologies for the diagnosis of atrial fibrillation: a systematic review and meta-analysis. *J Clin Med.* (2023) 12(20):6576. doi: 10.3390/jcm12206576

24. Lee S, Chu Y, Ryu J, Park YJ, Yang S, Koh SB. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. *Yonsei Med J.* (2022) 63(Suppl):S93–107. doi: 10.3349/ymj.2022.63.S93

25. Xie C, Wang Z, Yang C, Liu J, Liang H. Machine learning for detecting atrial fibrillation from ECGs: systematic review and meta-analysis. *Rev Cardiovasc Med.* (2024) 25(1):8. doi: 10.31083/j.rcm2501008

26. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521(7553):436–44. doi: 10.1038/nature14539

27. Deevi SA, Kaniraja CP, Mani VD, Mishra D, Ummer S, Satheesh C. HeartNetEC: a deep representation learning approach for ECG beat classification. *Biomed Eng Lett.* (2021) 11(1):69–84. doi: 10.1007/s13534-021-00184-x

28. Chang KC, Hsieh PH, Wu MY, Wang YC, Chen JY, Tsai FJ, et al. Usefulness of machine learning-based detection and classification of cardiac arrhythmias with 12-lead electrocardiograms. *Can J Cardiol.* (2021) 37(1):94–104. doi: 10.1016/j.cjca.2020.02.096

29. Bou Ezzeddine H, Vachulova A, Svetlosak M, Urban L, Hlivak P, Margitfalvi P, et al. Occurrence of symptoms after catheter ablation of atrial fibrillation. *BLL.* (2015) 116(08):461–4. doi: 10.4149/BLL_2015_086



OPEN ACCESS

EDITED BY

Rui Providencia,
University College London, United Kingdom

REVIEWED BY

Florian Straube,
Cardiology and Internal Intensive Care
Medicine, Germany
Anil Rajendra,
Grandview Medical Center, United States

*CORRESPONDENCE

G. André Ng
✉ andre.ng@leicester.ac.uk

RECEIVED 19 March 2025

ACCEPTED 06 June 2025

PUBLISHED 24 June 2025

CITATION

Antoun I, Abdelrazik A, Eldesouky M, Li X,
Layton GR, Zakkar M, Somani R and Ng GA
(2025) Artificial intelligence in atrial fibrillation:
emerging applications, research directions and
ethical considerations.
Front. Cardiovasc. Med. 12:1596574.
doi: 10.3389/fcvm.2025.1596574

COPYRIGHT

© 2025 Antoun, Abdelrazik, Eldesouky, Li,
Layton, Zakkar, Somani and Ng. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Artificial intelligence in atrial fibrillation: emerging applications, research directions and ethical considerations

Ibrahim Antoun^{1,2}, Ahmed Abdelrazik¹, Mahmoud Eldesouky¹,
Xin Li², Georgia R. Layton^{2,3,4}, Mustafa Zakkar^{2,4,5}, Riyaz Somani^{1,2}
and G. André Ng^{1,2,5*}

¹Department of Cardiology, Glenfield Hospital, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom, ²Department of Cardiovascular Sciences, Clinical Science Wing, Glenfield Hospital, University of Leicester, Leicester, United Kingdom, ³Department of Cardiac Surgery, Glenfield Hospital, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom, ⁴Department of Research, National Institute for Health Research Leicester Research Biomedical Centre, Leicester, United Kingdom, ⁵Leicester British Heart Foundation Centre of Research Excellence, Glenfield Hospital, Leicester, United Kingdom

Atrial fibrillation (AF) is the most prevalent sustained arrhythmia and a major contributor to stroke and heart failure. Despite progress in management, challenges persist in early detection, risk stratification, and personalised treatment. Artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), has emerged as a transformative tool in AF care. This scoping review examines the applications of AI across key domains: detection, risk prediction, treatment optimisation, and remote monitoring. AI-driven models enhance AF detection by analysing ECGs and wearable device data with high accuracy, enabling early identification of asymptomatic cases. By incorporating diverse clinical, imaging, and genomic data, predictive models outperform conventional risk scores in estimating stroke risk and disease progression. In treatment, AI assists in personalised anticoagulation decisions, catheter ablation planning, and optimising antiarrhythmic drug selection. Furthermore, AI-powered remote monitoring integrates wearable-derived insights with real-time decision support, improving patient engagement and adherence. Despite these advances, significant challenges persist, including algorithm transparency, bias, data integration, and regulatory hurdles. Explainable AI (XAI) is crucial to ensure clinician trust and facilitate implementation into clinical workflows. Future research should focus on large-scale validation, multi-modal data integration, and real-world AI deployment in AF management. AI has the potential to revolutionise AF care, shifting from reactive treatment to proactive, personalised management. Addressing current limitations through interdisciplinary collaboration will be key to realising AI's full potential in clinical practice and improving patient outcomes.

KEYWORDS

atrial fibrillation, artificial intelligence, machine learning, ECG, risk stratification, remote monitoring, personalised medicine

Introduction

Atrial fibrillation (AF) is the most common sustained arrhythmia worldwide, posing significant public health and clinical challenges due to its associated risks of stroke, heart failure, and mortality (1–3). AF management can be challenging, especially in the developing world, which has limited resources (4–11). However, AF is often intermittent and asymptomatic, making timely diagnosis challenging (12).

Artificial intelligence (AI) is the simulation of human intelligence in computers or machines, enabling them to perform tasks that typically require cognitive functions such as learning, reasoning, and problem-solving (13). Within AI, machine learning (ML) denotes a subset where algorithms learn from data to make predictions or decisions without being explicitly programmed (14). For example, ML can identify electrocardiogram (ECG) data patterns to distinguish between normal rhythm and atrial fibrillation. A more specialised subset of ML is deep learning (DL), which utilises layered neural networks inspired by the human brain. These models can process complex and high-dimensional data, such as continuous ECG signals or imaging data, and are particularly adept at uncovering subtle features that may be imperceptible to human clinicians.

AI techniques are increasingly employed in AF care to improve diagnostic accuracy, risk stratification, and treatment personalisation. For instance, DL algorithms have been shown to detect paroxysmal AF from sinus rhythm ECGs by learning subclinical signatures invisible to traditional analyses. Thus, a foundational understanding of these AI categories is essential for appreciating their applications in AF management, as explored throughout this review.

Over the last decade, AI has gained considerable momentum and is quickly becoming a mature discipline (15, 16). McCarthy coined the term AI in the late 1950s to denote the simulation of human intelligence in machines (17). Therefore, AI is not necessarily a newcomer, although most of its recent popularity is due to machine learning (ML). ML is a branch of AI that develops algorithms that use data to make predictions and improve their accuracy without being explicitly programmed to do so (18).

AI—particularly ML and DL techniques – have shown promise in improving AF detection, risk assessment, and management (19, 20) in recent years. Researchers are exploring AI across the spectrum of AF care, from early diagnosis using ECGs and wearables to personalised treatment selection. Despite encouraging results, there remain significant gaps in the literature and barriers to clinical implementation. Due to substantial advantages in big data processing, the use of AI in cardiovascular fields has recently aroused much attention. The use of AI in AF research has also increased significantly since 2012 (21).

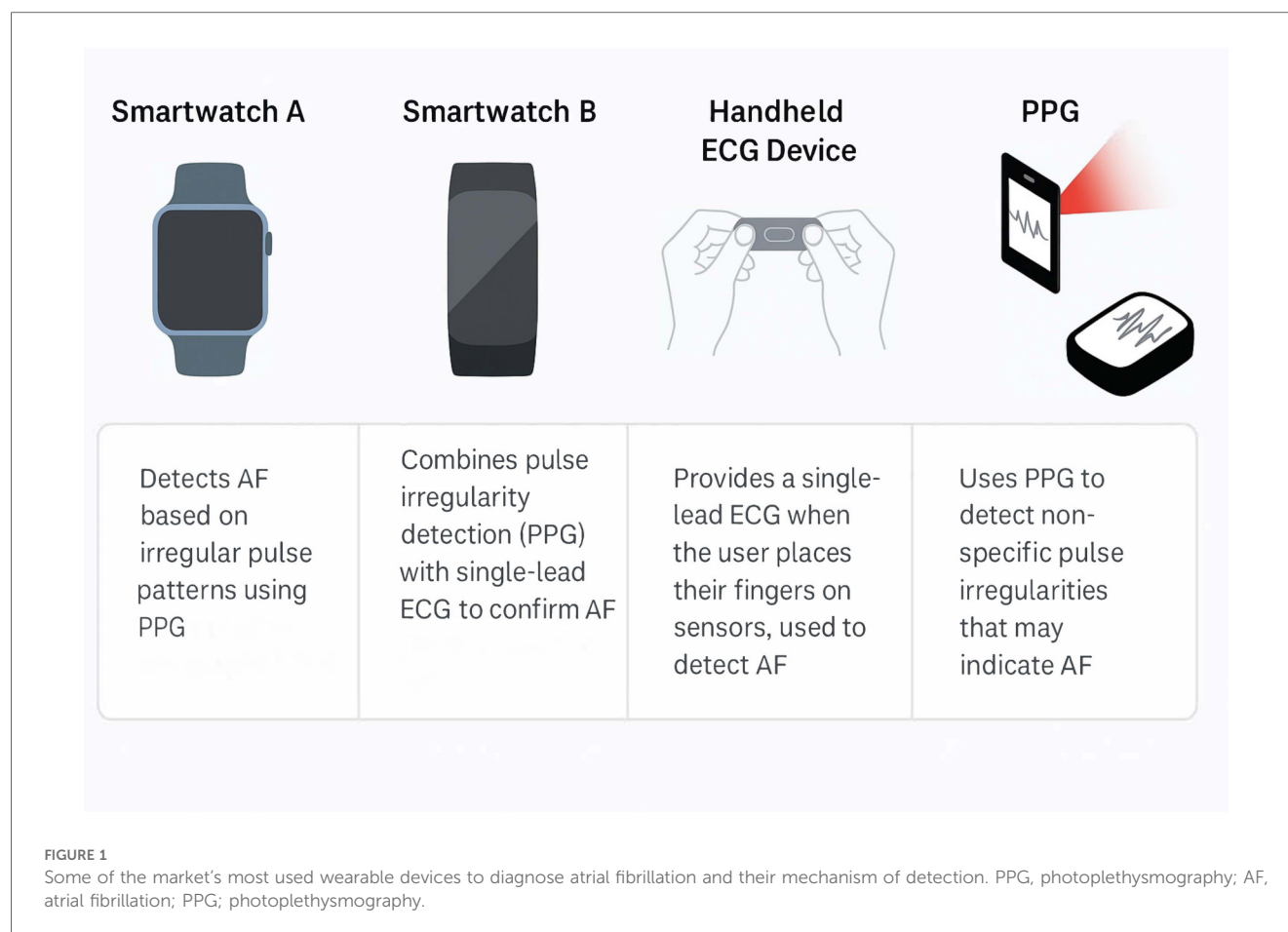
The following article provides a structured overview of key research areas, highlighting current advances, unmet needs, and potential methodologies for future exploration.

AI-driven early detection and diagnosis

AI algorithms can greatly enhance the early detection of AF by analysing large volumes of heart rhythm data from ECGs and wearable devices (22). Some currently available wearable devices are demonstrated in Figure 1. A traditional while the 12-lead ECG is considered the diagnostic gold standard for confirming AF, its sensitivity is limited in detecting paroxysmal AF (23). This is because the 12-lead ECG captures only a brief moment of cardiac electrical activity; thus, if the arrhythmia is not active during the recording, it may be missed. In contrast, wearable devices and continuous monitors can provide longer-duration rhythm surveillance, increasing the likelihood of detecting intermittent or asymptomatic episodes. ML models have shown high accuracy in detecting AF from single-lead ECGs or photoplethysmography (PPG) signals. For example, deep neural networks have achieved sensitivities and specificities in the 90%–99% range for classifying AF vs. normal rhythm using wearable ECG or PPG inputs (24, 25). Smartwatches equipped with FDA-cleared AF detection algorithms (using PPG and occasional ECG recordings) are increasingly popular and can reliably identify irregular pulse rhythms consistent with AF (26, 27). Notably, one smartphone-based PPG algorithm showed ~89% sensitivity and ~99% specificity compared to ECG diagnosis (28), highlighting the potential of ubiquitous devices for screening. This was also supported by a recent meta-analysis showing sensitivity of 92% and specificity of 96% in detecting AF on a single-lead ECG (29). AI can also detect subtle patterns in normal sinus ECGs that predict AF onset – in one study, an AI model predicted AF up to 4 h before an episode with an area under the curve (AUC) of 0.94. Another study of 180,922 patients showed that an AI-enabled ECG taken in normal sinus rhythm allows identification at the point of care of patients with AF (30).

Furthermore, the AI model predicts future episodes of AF with AUC 0.79 (0.72–0.86) in 122,394 patients, achieving the best prediction performance for males older than 70 years (31). These advances enable the detection of silent AF that would otherwise go unnoticed, allowing earlier interventions (e.g., starting anticoagulation to prevent stroke in asymptomatic patients) (32, 33). Despite promising accuracy, the real-world implementation of AI-driven AF detection faces challenges. Wearable-based algorithms can produce false positives or inconclusive alerts that require confirmatory traditional ECG (34–36). Many patients identified through smartwatch screening do not follow up with a physician, limiting clinical impact (37). Data from large digital health studies (Apple Heart Study and Huawei Heart Study) suggest that while technology can detect AF and even reduce stroke or hospitalisation rates when accompanied by proper follow-up, ensuring adherence to alerts by patients and providers is challenging (37, 38). There is also variability in performance across different devices and patient populations due to data quality and algorithm differences.

Future research may improve signal processing and ML robustness to reduce false alarms (for example, using ensemble models that combine ECG, PPG, and accelerometer data for



contextual intelligence). Unsupervised learning might uncover new AF-related patterns (e.g., in heart rate variability) that aid earlier detection (39). Moreover, integrating wearable data into clinical workflows via secure health information systems can ensure that clinicians are notified of significant events (40). Developing patient engagement strategies (adaptive app notifications or health coaching chatbots) to encourage prompt action after an AF alert will also be crucial to translating early detection into improved outcomes. While early detection of AF through wearable or implantable devices offers clear benefits, such as stroke prevention and timely intervention, it also raises concerns about overdiagnosis and overtreatment. The number needed to screen (NNS) to prevent one stroke varies depending on the population risk profile and screening method, with estimates ranging from 83 to 300 in moderate-risk populations. However, this remains an area of ongoing investigation (41).

Moreover, treating every detected AF episode with anticoagulation, especially short, asymptomatic episodes, may not be warranted. Recent evidence from the NOAH-AFNET 6 and ARTESiA trials suggests that oral anticoagulation in patients with device-detected subclinical AF may not significantly reduce stroke risk and might increase bleeding risk, particularly in patients without other high-risk features (42, 43). These findings highlight that not all AF episodes carry equal clinical risk, and treatment decisions should consider AF burden, episode duration, comorbidities, and patient-specific stroke risk.

As such, integrating AI-powered monitoring must complement clinician oversight and risk-based interpretation rather than automatically escalate to treatment. Future AI tools may help stratify the most clinically relevant episodes, aligning screening with personalised therapeutic thresholds. Many AI models using wavelets and ECG signals have shown high classification performance in AF detection, often exceeding 95% accuracy. These models vary in complexity and validation methods, as summarised in Table 1. High-quality studies, including large prospective cohorts and randomised trials assessing wearable-based AF detection, are summarised in Table 2.

AI for risk stratification and prediction of stroke or complications

Risk stratification is vital in AF to identify high-risk stroke, heart failure, or disease progression patients. Traditional clinical risk scores like CHA₂DS₂-VA are widely used for estimating stroke risk, but they have only modest discriminatory ability (C-statistic around 0.60) and ignore potentially important factors (70, 71). AI-based models offer a more individualised approach. Machine learning algorithms can ingest a wide range of clinical features – including demographics, comorbidities, lab results, imaging findings, and even social determinants of health – to predict adverse outcomes. Studies have shown that ML models

TABLE 1 Performance of selected major AF classification studies using wavelets and AI on ECG.

Study and year	Year	AI model	Evaluation method	Accuracy
Xu et al., (44)	2021	SVM-KNN-DT-ELM	Fold cross validation	98.6%
Singh et al., (45)	2020	SVM-LSTM	Fold cross validation	99.4%
Abdullah et al. (46)	2020	CNN-LSTM	Fold cross validation	98.1%
Ullah et al. (47)	2020	2-D CNN	Fold cross validation	99.1%
Zhao et al. (48)	2020	DCNN	Fold cross validation	87.8%
Nurmaini et al. (49)	2020	CNN-RNN-DNNs	Fold cross validation	99.1%
Kora et al. (50)	2019	SVM-KNN	Not published	94%-99.5%
Chashmi et al. (51)	2019	NN-SVM	10-fold cross validation	99%
Alarsan et al. (52)	2019	DT-RF-GDB	Fold cross validation	98%
Anwar et al. (53)	2018	NN	3-fold cross validation	99.8%
Lassoued et al. (54)	2017	ANN-MLP and neurofuzzy	GD + Momentum	99%
Xin et al. (55)	2017	SVM	Not published	92%
Filos et al. (56)	2017	SVM	Not published	93.8%
Kaya et al. (57)	2017	DT-SVM-NN	Fold cross validation	98.9%-99.3%
Saraswat et al. (58)	2017	PNN	Fold cross validation	100%
Dewangan et al. (59)	2016	ANN	Not published	87%
Thomas et al. (60)	2015	ANN	Fold cross validation	94.6%
Barmase et al. (61)	2013	Markov	Fold cross validation	99.8%
Sarkaleh et al. (62)	2012	MLP-NN	Fold cross validation	96.5%
Kim et al. (63)	2011	ELM	Not published	97.9%

SVM, support vector machines; KNN, K-nearest neighbour; DT, decision trees; ELM, extreme learning machine; LSTM, long short-term memory; CNN, convolutional neural network; RNN, recurrent neural network; DNN, deep neural network; NN, neural network; RF, random forests; GBD, gradient-boosted trees; ANN, artificial neural network; MLP, multilayer perceptron; PNN, probabilistic neural network.

TABLE 2 Key evidence from high-quality studies on wearables for AF detection.

Study name and year	Study type	Device/technology	Sample size	Main findings
Apple Heart Study, (64)	Prospective cohort study	Apple Watch (PPG-based)	419,297	PPV: 84% for irregular rhythm notification indicating AF
Huawei Heart Study, (38)	Prospective cohort study	Huawei Watch (PPG-based)	246,541	PPV of irregular pulse notifications: 91.6%; confirmed AF in 87%
mSToPS Trial, (65)	RCT	Zio XT patch (ECG-based)	2,659	AF newly diagnosed in 6.3% of actively monitored vs. 2.3% in controls
REHEARSE-AF Trial, (66)	RCT	AliveCor Kardia (single-lead ECG)	1,001	AF detection significantly increased (3.8% vs. 1%) using wearable vs. routine care
SCREEN-AF Trial, (67)	RCT	AliveCor KardiaMobile ECG	856 aged ≥ years old	Increased AF detection in high-risk elderly patients (5.3% vs. 0.5% in routine care)
Fitbit Heart Study, (68)	Prospective cohort study	Fitbit PPG	455,699	High accuracy (PPV: 98%) for AF detection in individuals with irregular heart rhythm alerts
Huawei heart study, (69)	Prospective cohort study	Huawei PPG	246,541	PPV: 91.6%, Both suspected AF and identified AF markedly increased with age

AF, atrial fibrillation; ECG, electrocardiogram; PPG, photoplethysmography; PPV, positive predictive value; RCT, randomised controlled trial.

outperform conventional scoring. For example, a convolutional neural network model achieved an AUC of 0.70 for predicting near-term stroke in AF patients, significantly better than the <0.50 AUC of CHA₂DS₂-VA in the same cohort. Researchers note that current scores are “rudimentary”, AI can capture additional risk predictors (like chronic kidney disease, prior heart failure, or socio-environmental factors) that improve accuracy (72, 73). Likewise, ML has been applied to predict other complications: one report used an AI algorithm on clinical and imaging data to identify patients likely to progress from paroxysmal to persistent AF or to develop heart failure (19). Another model could predict postoperative AF (a common surgical complication) with ~87% sensitivity and 83% specificity, enabling prophylactic strategies (74).

Stroke prediction and other outcomes

A major focus is AI use for stroke risk prediction in AF beyond standard scoring. ML models have been trained on extensive electronic health record datasets to calculate stroke risk automatically; these models can incorporate nuanced patterns (e.g., burden of AF, patterns of blood pressure control, or brain imaging markers) that static clinical scores cannot (75). Early results are encouraging, but validation is still required. Similarly, AI models have demonstrated the ability to predict AF recurrence after catheter ablation, guiding follow-up intensity and additional therapy. For instance, a DL algorithm using procedural data and patient characteristics predicted post-ablation AF recurrence with approximately 85% accuracy (76).

Including advanced phenotypic data like left atrial fibrosis on magnetic resonance imaging (MRI) or anatomical features on cardiac computed tomography (CT) is a promising avenue to further refine such predictions (77).

While many AI risk models have been developed, prospective validation and clinical uptake are key gaps. Few have been tested in randomised trials to show improved patient outcomes or cost-effectiveness (19). Future research should emphasise external validation of AI risk tools across diverse populations to ensure generalizability. Creating large, standardised, and labeled datasets (potentially through multi-center or federated learning collaborations) will help address current data heterogeneity issues (78). Moreover, researchers are exploring explainable AI techniques to identify which risk factors drive predictions, which could reveal novel modifiable risk factors (for example, an explainable model highlighted how combinations of social and clinical factors influence stroke risk in AF (79, 80)). Such insights might inform more holistic risk management strategies. In summary, AI-driven risk stratification promises to move beyond one-size-fits-all metrics toward nuanced, precision risk profiles for each AF patient.

AI-assisted treatment optimization (medications and ablation planning)

Managing AF involves choices between rate vs. rhythm control, various antiarrhythmic drugs, anticoagulation strategies, and interventional therapies like catheter ablation. AI can assist clinicians in tailoring these decisions to individual patients. One emerging application uses ML to determine which patients benefit most from a given therapy. For example, an AI-based causal forest model was recently developed using data from over 700,000 AF patients to personalise stroke prevention therapy (81). This model analyses patient characteristics to decide whether a patient would have better outcomes with lifelong anticoagulation (using a direct oral anticoagulant, DOAC) or a left atrial appendage occlusion (LAAO) procedure. Notably, it can identify subgroups of patients for whom LAAO provides a positive net benefit (reducing stroke risk without undue bleeding risk) vs. those who fare better on medication (81). Such AI-driven decision support could fill an important gap in current practice, where selecting candidates for LAAO vs. DOAC is often unclear. Similarly, ML has been applied to recommend optimal rate or rhythm control strategy by predicting outcomes like symptom improvement or hospitalisation risk under each approach (82). However, more research is needed in this area. A significant gap is the lack of clinical trial evidence demonstrating that AI-guided treatment decisions improve long-term outcomes in AF. Ongoing studies will reveal the real-world impact, particularly in applying AI recommendations for therapy selection in a prospective trial. Methodologically, incorporating reinforcement learning could prove beneficial—an AI that “learns” the optimal treatment through trial and error on patient data might suggest dynamic treatment adjustments, such as escalating from drugs to ablation if specific patterns arise. Furthermore, integrating AI into electronic health records as a

clinical decision support tool at the point of care is another avenue to explore, but it must be implemented in a way that aligns with the clinician’s workflow and provides transparent reasoning for recommendations (83). Combining an AI’s predictive power with a physician’s clinical judgment may yield the best outcomes in tailoring AF therapy plans.

AI role in catheter ablation

Catheter ablation is an effective rhythm-control therapy; success rates can differ, and repeat procedures are common (84). AI has the potential to enhance patient selection and procedural planning for ablation. An innovative study employed DL on cardiac imaging data to predict non-pulmonary vein (PV) triggers of AF before an ablation procedure (85). Typically, ablation targets triggers in the PVs, but patients with additional atypical trigger sites often experience recurrence. The AI model correctly predicted these non-PV trigger locations in ~82% of cases (64% sensitivity, 88% specificity), improving the overall accuracy of identifying all trigger sites to 89% (85). This information can help electrophysiologists personalise the ablation strategy rather than using a uniform approach for all patients. AI has also been utilised in intra-procedural mapping – for instance, algorithms that rapidly interpret electrogram patterns to distinguish AF drivers or to titrate energy delivery. Early clinical experience with AI-guided ablation dosing, including high-power short-duration ablation with algorithmic monitoring for safety, shows the potential to reduce complications (86). Additionally, AI can assist in medication optimisation by predicting an individual’s response or side-effect risk to a particular antiarrhythmic drug based on their profile, though this is still largely theoretical (87). Recent high-quality evidence from the TAILORED-AF trial supports using AI-guided ablation in persistent AF. This randomised, double-blind trial showed that targeting AI-identified spatio-temporal electrogram dispersion areas in addition to standard PVI significantly improved 12-month AF freedom rates (88% vs. 70%, $P < 0.0001$). While safety was comparable, procedure duration was longer. These results validate AI’s role in refining ablation strategy but highlight the need for further trials to assess long-term outcomes, reproducibility, and workflow integration (88). New AI-based software solutions were designed to assist operators in targeting AF drivers. Acute and long-term outcomes suggest that the AI-based AF electrogram software delivers simple perioperative cues, ensuring standardisation across multiple platforms, catheters, and operators (89). For example, a recent study demonstrated that DISPERS-guided ablation using ML software (the Volta VX1 software) and PVI for long-standing persistent AF caused a lower risk of AF recurrence in long-term follow-ups (90).

Machine learning models for personalized AF management

Personalised medicine in AF aims to move beyond generalised treatment guidelines and towards individualised care plans. AI is a

key enabler of this vision, as it can analyse each patient's unique combination of factors. Holistic ML models can assimilate diverse data, including genomics, biomarkers, lifestyle factors, and detailed disease history, to define patient subgroups or “phenotypes” of AF. For instance, clustering algorithms (unsupervised ML) have been used to identify novel AF phenotypes that might respond differently to treatments (39). For example, a patient might have AF driven largely by obesity and hypertension (risk-factor-mediated AF) (91). At the same time, another's AF might be linked to specific genetic variants or fibrotic scar burden in the atrium. AI can help classify such subgroups, which is the first step to personalised therapy, including aggressive risk factor modification for one phenotype vs. early ablation for another.

Early rhythm control has been proven beneficial as AF begets AF, and early intervention has shown better results irrespective of the mechanisms. As there is an evolution of AF in many cases (triggers at the beginning with short episodes and fibrosis with persistent AF types), AI might help identify patients at risk for AF and help establish a primary preventive therapy. In case of AF, early treatment should be offered irrespective of risk factors, as e.g., those with heart failure benefit most. Researchers have proposed new AF classifications using ML, which are being studied for their prognostic and therapeutic relevance (92). An example of such translational work is the ARISTOTELES project, which uses AI to integrate clinical, imaging, genetic, and biomarker data to personalise risk prediction and treatment in patients with AF and multimorbidity. The project aims to refine stroke and bleeding risk stratification in the context of oral anticoagulation and to address therapeutic decision-making complexities in patients with coexisting conditions such as heart failure, diabetes, and chronic kidney disease. By embracing a holistic, data-driven approach, ARISTOTELES supports guideline-aligned, individualised care and seeks to improve outcomes, reduce adverse events, and optimise resource use in real-world clinical settings (93). Key studies exploring AI-based classification and prediction of AF, particularly those using ECG or clinical data for risk stratification, are outlined in Table 3.

Integration of AI with genomic and clinical data

In the era of precision medicine, merging AI with genomics presents a promising methodology. Extensive genome-wide association studies have identified numerous genetic loci associated with AF, yet interpreting these for individual risk remains complex. AI can bridge this gap by integrating genetic risk scores with phenotypic data. For instance, one analysis demonstrated that AI-driven ECG analysis serves as a practical and cost-effective means of predicting AF risk and onset, capturing lifetime cardiac variations, whereas genomics offers a more static risk profile; the combination of the two facilitates “truly individualised care” that transcends the average patient model (96).

This might mean an AI model uses a patient's ECG and blood biomarkers to detect subtle signs of atrial remodelling while incorporating their genetic predisposition to refine risk and guide preemptive therapy (97). However, a meta-analysis indicated that AF prediction using AI is still underdeveloped, though DL techniques are becoming increasingly accurate. Nevertheless, these methods are not being applied as frequently as expected (98).

Challenges and research opportunities

Achieving personalised AF management with AI faces several hurdles. One is data silos – the need to gather comprehensive datasets that include outcomes of different management strategies in diverse patient profiles. Collaborative consortia and data-sharing with privacy protections could help amass enough data for robust personalised models. Another challenge is interpretability: Clinicians will require understandable explanations for why an AI recommends a personalised approach. Research into explainable AI for personalised medicine is, therefore, critical (99, 100). Furthermore, prospective trials are needed to test AI-guided personalised management: for example, an algorithm might propose varying treatment strategies for

TABLE 3 Main evidence on AI-assisted treatment in atrial fibrillation.

Study/Author (year)	AI Application	Population	Main findings
Ngufo et al. (81)	Personalized anticoagulation vs. LAAO	744,190 AF patients	AI model identifies optimal stroke prevention therapy, highlighting patients benefiting most from DOAC or LAAO
Kim et al. (94)	AI use in guiding rhythm management	≈ 42,000 AF patients	Healthcare systems using algorithms for AF rhythm management must balance prediction accuracy with model interpretability
Liu et al. (85)	Catheter ablation strategy	521 patients undergoing PAF ablation	DL predicted non-PV triggers with 82% accuracy (88% specificity, 64% sensitivity)
Deisenhofer et al. (TAILORED-AF Trial) (95)	AI-guided catheter ablation	370 persistent AF patients (AI tailored arm, $n = 187$)	AI-guided ablation plus standard PVI improved 12-month AF-free rates (88% vs. 70%, $p < 0.0001$)
Bahlke et al. (90)	ML-guided DISPERS ablation software	50 persistent AF undergoing ablation	ML-assisted ablation software reduced long-term recurrence rates compared to standard PVI
Sanchez de la Nava et al. (87)	Antiarrhythmic drug selection	127 AF patient models	AI predicted patient-specific drug responses in silico
Seitz et al. (89)	Standardization of ablation outcomes	85 persistent AF patients	AI software standardized electrogram-based ablation across multiple operators and platforms

AF, atrial fibrillation; AI, artificial intelligence; DL, deep learning; DOAC, direct oral anticoagulants; LAAO, left atrial appendage occlusion; ML, machine learning; PVI, pulmonary vein isolation; PV, pulmonary vein; PAF, paroxysmal atrial fibrillation.

patients based on their cluster phenotype – testing this against usual care will show if personalisation via AI improves outcomes like AF recurrence or quality of life. If successful, these approaches could pave the way to truly precision cardiology, where every AF patient's management is dynamically tailored by AI insights drawn from patients “like them” in large databases.

The role of AI in remote monitoring and patient adherence

Remote monitoring technologies for AF allow continuous or frequent rhythm surveillance outside the clinic. AI plays a vital role in interpreting the large volume of data generated by remote monitoring devices. Smartwatches and patches can detect arrhythmias in real time, but AI algorithms must distinguish true AF episodes from noise or benign irregularities (101). When deployed effectively, AI-driven remote monitoring can alert clinicians to AF onset or recurrence, enabling earlier intervention (68). Studies have demonstrated that mobile health interventions for AF can reduce healthcare utilisation combined with algorithmic monitoring. In one cluster trial, patients supported by a mobile app and wearables had significantly lower rehospitalisation and adverse event rates than those with standard care (38). Another study showed that contactless AI monitoring could accurately detect AF without any wired device (102, 103). This sets the stage for futuristic remote surveillance methods that are seamless for patients.

A critical aspect of remote AF management is ensuring patients adhere to monitoring and therapy. AI can assist here through personalised feedback and coaching. For example, smartphone apps with a conversational “relational agent” have been piloted to engage AF patients daily, provide education, and encourage medication adherence. In a 120-patient trial, those randomised to a 30-day smartphone app with an AI-driven virtual coach and a portable ECG monitor showed significant improvements in adherence and quality of life compared to controls (104). This suggests that AI can help close the gap between detecting AF and prompting patients to act, including taking medications and contacting healthcare providers. Nonetheless, challenges remain: Large-scale screening studies found that some people ignored or delayed responding to AF alerts on their devices (37). This highlights that technology alone is insufficient; behavioral science must be integrated into AI systems.

Future research should explore adaptive notification systems that adjust the urgency and style of alerts based on patient behavior patterns to avoid alarm fatigue while conveying importance (105). AI might predict which patients are at risk of non-adherence. This can be done by analysing their past application usage, heart rate trends, or speech patterns in consultations (106). Combining remote monitoring AI with telemedicine services is another promising avenue. If an algorithm detects AF, it could automatically schedule a telehealth visit or message a healthcare provider, streamlining the response. AI can turn passive remote monitoring into an active, responsive system that detects AF, facilitates prompt management, and keeps patients engaged in their care.

Explainability and ethical concerns in AI-based AF detection and treatment

Many AI models, especially DL, are often criticised as “black boxes” – they make predictions (AF detected or stroke risk high) without an easily interpretable rationale. In the context of AF, lack of explainability can hinder clinician trust and adoption. Researchers have started integrating explainable AI (XAI) techniques into their models to address this. For instance, one study converted PPG pulse data into images and used a convolutional neural network to classify AF; importantly, they incorporated XAI methods to highlight which signal features contributed to the classification, providing transparency to clinicians (107). The resulting model achieved 100% accuracy in distinguishing AF from normal rhythm while ensuring the decision process was interpretable. Such approaches allow physicians to verify that the AI detects physiologically relevant patterns (like irregular RR intervals or fibrillatory waves) rather than spurious noise. Explainability is equally crucial in AI-driven treatment recommendations – doctors need to understand why an algorithm favors a particular therapy for a patient (perhaps due to that patient's combination of age, stroke risk, and prior haemorrhage history) to feel comfortable following the advice (108). Developing user-friendly visualisation tools and explanation summaries for AI outputs is an active area of research that will make AI more ethically and clinically palatable.

The use of AI in AF raises several ethical considerations. Patient data privacy is paramount, as AI models often require large datasets (ECGs, wearable records, and health records) that may contain sensitive information. Ensuring compliance with privacy regulations and using data anonymisation or federated learning (where data stay at hospital sites and only model updates are shared) can mitigate privacy risks (73). Bias is another concern – if an AI model is trained mostly on certain demographics, it may perform less accurately for underrepresented groups, potentially exacerbating healthcare disparities. For example, an algorithm trained predominantly on younger patients might miss AF in the elderly or vice versa. Researchers have pointed out that algorithms must consider social determinants of health and diverse patient attributes to avoid bias (109). Ongoing efforts to use diverse training datasets and to audit algorithms for fairness are critical (110). Additionally, there is the ethical question of handling false positives/negatives: a false positive AF alert can cause anxiety and unnecessary testing, while a false negative might give false reassurance. Striking the right balance in algorithm sensitivity is partly a clinical value judgment. Some ethicists have raised concerns about widespread consumer AF screening being promoted without clear guidance, potentially putting users at risk of over-treatment or anxiety for the sake of tech company marketing (111, 112).

Several strategies are recommended to ensure that AI in AF is used responsibly. First, ethicists and patient representatives should be involved early in developing AI tools to identify concerns such as consent and data ownership. Second, incorporate the core principles of biomedical ethics: beneficence (the AI should demonstrably assist patients), nonmaleficence (minimising harm from errors), autonomy (patients should control how their data

are used and be informed about AI's involvement in their care), and justice (equitable access to the benefits of AI) (113, 114). Specifically, this could mean providing patients with a straightforward opt-in/out option for data sharing and ensuring that AI tools are accessible in community hospitals, not just academic centres. Third, maintain a human-in-the-loop approach: AI should support, rather than replace, clinician decision-making, and clinicians should override or question AI when it conflicts with clinical judgment or patient preferences (115). The medical community can harness their advantages while upholding high ethical standards by making AI systems transparent, secure, and patient-centred. AI-driven precision medicine in AF must move beyond purely clinical or algorithmic outputs to incorporate patients' own perceptions of safety and autonomy—factors deeply influenced by cultural background, previous healthcare experiences, and psychosocial context. Overreliance on binary risk models or population-level predictions risks marginalising patient values and eroding trust. Ethically grounded AI must therefore embed principles of shared decision-making, allowing patients to weigh algorithmic recommendations against their personal goals and beliefs. This includes providing transparent explanations of AI outputs and fostering cultural competence in both data design and clinical implementation to support genuinely patient-centred care.

Challenges in implementing AI for AF in clinical practice

Despite the increasing research, there is a recognised gap between AI models developed in laboratories and the tools that physicians use at the bedside. One significant challenge is rigorous clinical validation (116). Many AI algorithms for AF detection or risk prediction have been tested retrospectively or on limited datasets (117, 118); few have been evaluated in prospective clinical trials or real-world practice settings. Without evidence that AI improves patient outcomes or workflow, healthcare providers may hesitate to adopt these tools. Regulatory approval pathways for AI in medicine are evolving – algorithms may require clearance as medical devices, and there are questions about how to regulate AI systems that continuously learn and update. Obtaining regulatory approval can be complex and time-consuming, particularly if an AI's decision-making logic is not easily interpretable to regulators.

Implementing AI in daily practice also presents logistical challenges. Hospitals must integrate AI software with electronic health record systems and device data streams, ensuring reliability and cybersecurity. Clinicians experience alert fatigue from existing monitoring systems; introducing AI alerts or recommendations could further burden them if not carefully designed (119). Therefore, human factors engineering is essential – AI tools must be intuitive, with concise and relevant outputs. Clinician training is another crucial aspect: cardiologists and general practitioners will require a fundamental understanding of how the AI operates and its limitations to use it effectively and maintain trust in the system (120). At the practice level, some resistance to new technology is

natural; early adopters must advocate for successful use cases to persuade their peers of AI's potential value. High-quality data is the fuel for AI, and data can often be messy in practice. AF-related data frequently resides in disparate sources and may lack standardisation. As noted in the review article by Popat et al., variability and heterogeneity in data and methods have led to the inconsistent performance of AI tools across studies (73). This suggests that an AI model may not generalise effectively without standard data formats and solid data governance when implemented in a different hospital or demographic. Initiatives such as establishing shared data repositories and adopting common standards for documenting arrhythmia data can be beneficial. Another obstacle is the computational infrastructure – not all clinics can execute advanced AI algorithms in real time. Cloud-based solutions could alleviate this issue; however, data security and latency concerns may occur.

To bridge these gaps, researchers and healthcare systems are beginning to collaborate on implementing science for AI. This includes pilot programmes where AI tools are introduced in a controlled manner, and their impact on decision-making, outcomes, and clinician workload is measured. Feedback from these pilots can guide iterative improvements. Additionally, clear guidelines from professional societies on how to incorporate AI into AF management (when to trust an AI-detected AF episode or how to use an AI risk score in anticoagulation decisions) will provide reassurance and standardisation (121). Addressing implementation challenges will ultimately require a multidisciplinary approach, with data scientists, information technology specialists, clinicians, and administrators working together to ensure that AI for AF is accurate, useful, and seamlessly embedded in care delivery.

Future directions for AI research in AF

The intersection of AI and AF management is a rapidly evolving field with several exciting avenues for future research to address current gaps. A clear need exists for developing standardised, well-annotated datasets for AF. Future research could focus on building large, shared databases of ECGs (including those in sinus rhythm and AF), patient outcomes, and imaging data, which would enable more robust model training and validation. International collaborations and data-sharing agreements, employing privacy-preserving techniques, will expedite this progress and minimise duplication of effort. Integrating various data modalities presents a promising research frontier. AF is a multi-factorial disease; thus, combining multi-modal data in AI models may yield new insights. For instance, researchers could develop models that input ECG signals, cardiac MRI scans, genetic information, lab results, and wearable activity logs to provide a comprehensive risk assessment or guide therapy. Initial efforts to combine ECG-based AI with genomics have demonstrated the potential for more precise predictions (96). Future studies will likely build on this by including proteomics or metabolomics to capture substrate changes in AF. Such comprehensive models could, for example, predict which

patients will respond to upstream therapies, such as aggressive risk factor management or anti-inflammatory treatment, based on their unique biomarker signature.

With the proliferation of wearable devices, one future direction is to deploy AI algorithms that operate in real time on device data. A smartwatch could detect AF and predict when an episode is likely to occur based on recent trends, alerting the patient to take preventive actions, such as performing a breathing exercise or taking a pill (122). Research into lightweight AI models that can run on wearable hardware or smartphones for continuous monitoring without relying on cloud servers will be crucial for responsive and scalable AF management. Future AI research may also contribute to the development of new treatments for AF. By analysing large datasets of drug responses, AI could aid drug discovery by identifying novel molecules or repurposing existing medications that might effectively treat AF. An AI-driven “autopilot” for PVI is on the horizon. These include algorithms capable of guiding catheters during ablation with minimal human input, optimising lesion placement and duration in real time based on patient-specific atrial anatomy and electrophysiology (89). Preliminary steps in this direction, such as AI mapping of electrical rotors or lesion assessment through imaging, warrant further investigation exploration. As AI systems become more complex, a parallel research priority is enhancing their explainability and user trust. Future AI models for AF might incorporate built-in explanation modules that translate their complex computations into cardiologist-friendly explanations, highlighting which part of a P wave or what clinical factor led to a high-risk prediction. Measuring and improving clinician trust in AI recommendations will likely become a research topic, possibly

by testing different explanation strategies in user studies. This also ties into medicolegal considerations – defining how responsibility is shared between AI tools and human providers – which may be informed by research in ethics and legal frameworks.

Finally, the ultimate test for any AI application is demonstrating improved patient outcomes. Future research must include prospective, randomised trials where AI-guided care is compared to standard care in AF. These trials could examine endpoints like reduced AF burden, hospital admissions, stroke rates, or improved quality of life. For example, a trial might randomise clinics to use an AI-guided AF screening and management protocol vs. usual guideline-based management to see if the AI arm achieves better outcomes. Additionally, health economics research should evaluate the cost-effectiveness of AI interventions. Positive results will be key to convincing guidelines committees and payers to endorse AI tools in routine AF care (19). A summary of AI use in diagnosing, classifying and treating AF is demonstrated in Figure 2. The main take-home messages from this review, covering diagnostic, therapeutic, ethical, and implementation aspects of AI in AF management, are summarised in Table 4.

Conclusion

In conclusion, AI holds immense potential to transform AF detection and management—from identifying the arrhythmia earlier and more accurately to personalising therapy decisions and continuously supporting patients in their daily management. The existing literature provides a robust foundation but also highlights gaps, such as inconsistent tool performance and

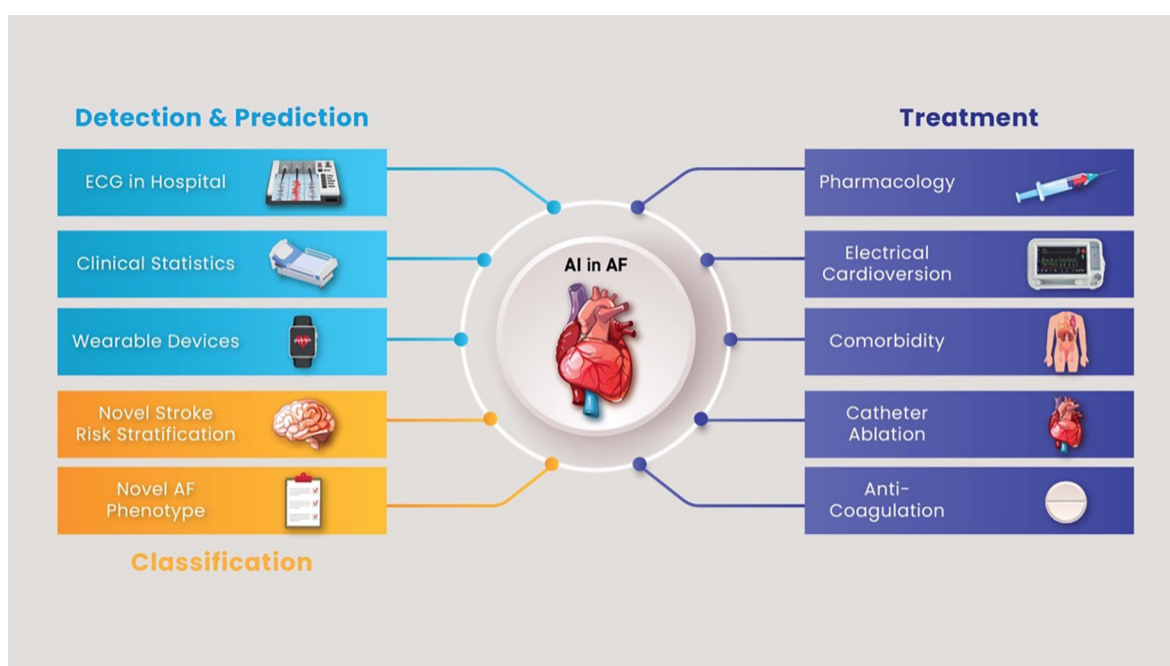


FIGURE 2
Summary of artificial use in the diagnosis and treatment of atrial fibrillation. AF, atrial fibrillation; AI, artificial intelligence; ECG, electrocardiogram.

TABLE 4 Main take-home messages of the article.

Domain	Key take-home messages
Early detection and diagnosis	AI significantly improves early detection of AF using ECG and wearable data, but false positives and clinical follow-up remain challenges.
Risk stratification and prediction	AI-driven predictive models outperform traditional risk scores, providing precise stroke and AF progression risk assessments.
Treatment optimization	AI supports personalized decisions in anticoagulation therapy, antiarrhythmic drug selection, and catheter ablation planning, improving patient outcomes.
Remote monitoring and adherence	AI-enabled remote monitoring enhances continuous patient surveillance and adherence yet necessitates behavioural interventions for optimal clinical impact.
Ethical and Explainability concerns	Explainable AI (XAI) techniques are essential to ensure transparency, clinician trust, and ethical implementation, reducing bias and promoting equitable care.
Challenges in clinical implementation	Clinical validation, regulatory approval, integration with clinical workflows, and clinician training are crucial barriers needing attention to enable widespread AI adoption.
Future research directions	Priorities include large-scale prospective validation trials, multimodal data integration, standardizing datasets, and leveraging AI for personalized, proactive AF management.

AF, atrial fibrillation; AI, artificial intelligence; ECG, electrocardiogram; XAI, explainable artificial intelligence.

limited clinical validation. By focusing on the identified research directions—enhancing data quality, ensuring ethical implementation, and rigorously testing AI in practice—the next wave of studies can help realise AI’s promise in atrial fibrillation. In the coming years, we will see AI progress from exploratory trials into integrated clinical practice, ultimately improving outcomes and quality of life for patients with AF, provided we address the challenges and learn from ongoing research at every step. Despite promising advancements across detection, risk stratification, treatment optimisation, and remote monitoring, the clinical implementation of AI in AF remains hindered by a persistent lack of large-scale prospective RCT evidence. This limitation, common across many AI applications in medicine, highlights the need for rigorous validation to establish real-world efficacy, safety, and cost-effectiveness. Addressing this evidence gap through well-designed, multi-centre RCTs will be essential for translating AI innovations into routine clinical practice.

Author contributions

IA: Writing – original draft, Investigation, Software, Conceptualization. AA: Writing – review & editing. ME: Writing – review & editing. XL: Writing – review & editing. GL: Writing – review & editing. MZ: Writing – review & editing. RS: Writing – review & editing. GN: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. G Andre Ng is

supported by British Heart Foundation Research Excellence Award (RE/24/130031), British Heart Foundation Programme Grant (RG/17/3/32774), Medical Research Council Biomedical Catalyst Developmental Pathway Funding Scheme (MR/S037306/1) and NIHR i4i grant (NIHR204553). Mustafa Zakkar is supported by British Heart Foundation award (CH/12/1/29419) to the University of Leicester and Leicester NIHR Biomedical Research Centre (NIHR203327).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Grammarly was used for proofreading the article

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham heart study: a cohort study. *Lancet*. (2015) 386(9989):154–62. doi: 10.1016/S0140-6736(14)61774-8

2. Lip GY, Kakar P, Watson T. Atrial fibrillation—the growing epidemic. *Heart*. (2007) 93(5):542. doi: 10.1136/hrt.2006.110791

3. Antoun I, Layton GR, Abdelrazik A, Eldesouky M, Zakkar M, Somani R, et al. The pathophysiology of sex differences in stroke risk and prevention in atrial fibrillation: a comprehensive review. *Medicina (Kaunas)*. (2025) 61(4):649. doi: 10.3390/medicina61040649

4. Antoun I, Alkhayer A, Aljabal M, Mahfoud Y, Alkhayer A, Simon P, et al. Thirty-day unplanned readmissions following hospitalization for atrial fibrillation in a tertiary

Syrian centre: a real-world observational cohort study. *Heart Rhythm O2*. (2024) 5:860–1. doi: 10.1016/j.hroo.2024.05.010

5. Antoun I, Aljabal M, Alkhayer A, Mahfoud Y, Alkhayer A, Simon P, et al. Atrial fibrillation inpatient management patterns and clinical outcomes during the conflict in Syria: an observational cohort study. *Perfusion*. (2024) 40:02676591241259140. doi: 10.1177/02676591241259140

6. Antoun I, Alkhayer A, Aljabal M, Mahfoud Y, Alkhayer A, Kotb AI, et al. Incidence, outcomes, and predictors of new heart failure in Syrian conflict-affected population following hospitalization for atrial fibrillation: a retrospective cohort study. *Perfusion*. (2024) 02676591241283883. doi: 10.1177/02676591241283883

7. Antoun I, Alkhayer A, Alkhayer A, Yazji K, Somani R, Ng GA, et al. Six-month emergent readmissions following hospitalization for atrial fibrillation amid the Syrian conflict: a real-world observational cohort study. *J Cardiovasc Electrophysiol*. (2025) 36:582–8. doi: 10.1111/jce.16579

8. Antoun I, Alkhayer A, Eldin AJ, Alkhayer A, Yazji K, Somani R, et al. Gender disparity in quality of life in atrial fibrillation patients during the Syrian conflict: an observational cohort study. *Heart Rhythm O2*. (2025) 6:362–7. doi: 10.1016/j.hroo.2024.12.010

9. Antoun I, Alkhayer A, Aljabal M, Alkhayer A, Simon P, Mahfoud Y, et al. The validity and reliability of the Arabic version of the EQ-5D in atrial fibrillation patients in a conflict country: a study from Syria. *BMC Cardiovasc Disord*. (2024) 24(1):541. doi: 10.1186/s12872-024-04203-4

10. Nguyen TN, Hilmer SN, Cumming RG. Review of epidemiology and management of atrial fibrillation in developing countries. *Int J Cardiol*. (2013) 167(6):2412–20. doi: 10.1016/j.ijcard.2013.01.184

11. Santos IS, Goulart AC, Olmos RD, Thomas GN, Lip GY, Lotufo PA, et al. Atrial fibrillation in low- and middle-income countries: a narrative review. *Eur Heart J Suppl*. (2020) 22(Supplement_O):O61–77. doi: 10.1093/eurheartj/suaa181

12. Boriani G, Laroche C, Diemberger I, Fantecchi E, Popescu MI, Rasmussen LH, et al. Asymptomatic atrial fibrillation: clinical correlates, management, and outcomes in the EORP-AF pilot general registry. *Am J Med*. (2015) 128(5):509–18.e2. doi: 10.1016/j.amjmed.2014.11.026

13. Holmes J, Sacchi L, Bellazzi R. Artificial intelligence in medicine. *Ann R Coll Surg Engl*. (2004) 86:334–8. doi: 10.1308/147870804290

14. Mahesh B. Machine learning algorithms-a review. *Int J Sci Res*. (2020) 9(1):381–6. doi: 10.21275/ART20203995

15. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. *Med Phys*. (2020) 47(5):e228–35. doi: 10.1002/mp.13562

16. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int J Inf Manage*. (2021) 57:101994. doi: 10.1016/j.ijinfomgt.2019.08.002

17. Russel S, Norvig P. Artificial intelligence—a modern approach 3rd edition. *Knowl Eng Rev*. (2012) 1:78–9.

18. Bishop CM, Nasrabadi NM. *Pattern Recognition and Machine Learning*. New York, NY: Springer (2006).

19. Liang H, Zhang H, Wang J, Shao X, Wu S, Lyu S, et al. The application of artificial intelligence in atrial fibrillation patients: from detection to treatment. *Rev Cardiovasc Med*. (2024) 25(7):257. doi: 10.31083/j.rcm2507257

20. Menezes Junior ADS, E Silva ALF, E Silva LRF, de Lima KBA, Oliveira HL. A scoping review of the use of artificial intelligence in the identification and diagnosis of atrial fibrillation. *J Pers Med*. (2024) 14(11):1069. doi: 10.3390/jpm14111069

21. Olier I, Ortega-Martorell S, Pieroni M, Lip GY. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. *Cardiovasc Res*. (2021) 117(7):1700–17. doi: 10.1093/cvr/cvab169

22. Abdelrazik A, Eldesouky M, Antoun I, Lau EYM, Koya A, Vali Z, et al. Wearable devices for arrhythmia detection: advancements and clinical implications. *Sensors*. (2025) 25(9):2848. doi: 10.3390/s25092848

23. Harris K, Edwards D, Mant J. How can we best detect atrial fibrillation? *J R Coll Physicians Edinb*. (2012) 42:5–22. doi: 10.4997/JRCPE.2012.S02

24. Playford D, Jais P, Weerasooriya R, Martyn S, Bollam L, Turewicz M, et al. A validation study of automated atrial fibrillation detection using alerte digital health's artificial intelligence system. *Heart Lung Circ*. (2017) 26:S279–80. doi: 10.1016/j.hlc.2017.06.546

25. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the apple heart study. *Am Heart J*. (2019) 207:66–75. doi: 10.1016/j.ahj.2018.09.002

26. Papalamprakopoulou Z, Stavropoulos D, Moustakidis S, Avgerinos D, Efremidis M, Kampaktis PN. Artificial intelligence-enabled atrial fibrillation detection using smartwatches: current status and future perspectives. *Front Cardiovasc Med*. (2024) 11:1432876. doi: 10.3389/fcvm.2024.1432876

27. Chen E, Jiang J, Su R, Gao M, Zhu S, Zhou J, et al. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. *Heart Rhythm*. (2020) 17(5):847–53. doi: 10.1016/j.hrthm.2020.01.034

28. Brasier N, Raichle CJ, Dörr M, Becke A, Nothmann V, Weber S, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). *Europace*. (2019) 21(1):41–7. doi: 10.1093/europace/euy176

29. Manetas-Stavarakakis N, Sotiropoulou IM, Paraskevas T, Maneta Stavarakaki S, Bampatsias D, Xanthopoulos A, et al. Accuracy of artificial intelligence-based technologies for the diagnosis of atrial fibrillation: a systematic review and meta-analysis. *J Clin Med*. (2023) 12(20):6576. doi: 10.3390/jcm12206576

30. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, 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):861–7. doi: 10.1016/S0140-6736(19)31721-0

31. Melzi P, Tolosana R, Cecconi A, Sanz-Garcia A, Ortega GJ, Jimenez-Borreguero LJ, et al. Analyzing artificial intelligence systems for the prediction of atrial fibrillation from sinus-rhythm ECGs including demographics and feature visualization. *Sci Rep*. (2021) 11(1):22786. doi: 10.1038/s41598-021-02179-1

32. Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, et al. Atrial fibrillation in patients with cryptogenic stroke. *N Engl J Med*. (2014) 370(26):2467–77. doi: 10.1056/NEJMoal311376

33. Hart RG, Diener H-C, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. (2014) 13(4):429–38. doi: 10.1016/S1474-4422(13)70310-7

34. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol*. (2019) 12(6):e006834. doi: 10.1161/CIRCEP.118.006834

35. Tran K-V, Filippaios A, Noorishirazi K, Ding E, Han D, Moghaghghian F, et al. False atrial fibrillation alerts from smartwatches are associated with decreased perceived physical well-being and confidence in chronic symptoms management. *Cardiol Cardiovasc Med*. (2023) 7(2):97. doi: 10.26502/fccm.92920314

36. Sepehri Shamloo A, Bollmann A, Dagnes N, Arya A, Hindricks G. Smart watch devices for atrial fibrillation screening: it has to start somewhere. *J Am Coll Cardiol*. (2020) 75(11):1364–5. doi: 10.1016/j.jacc.2019.10.063

37. Brandes A, Stavrakis S, Freedman B, Antoniou S, Boriani G, Camm AJ, et al. Consumer-led screening for atrial fibrillation: frontier review of the AF-SCREEN international collaboration. *Circulation*. (2022) 146(19):1461–74. doi: 10.1161/CIRCULATIONAHA.121.058911

38. Guo Y, Lane DA, Wang L, Zhang H, Wang H, Zhang W, et al. Mobile health technology to improve care for patients with atrial fibrillation. *J Am Coll Cardiol*. (2020) 75(13):1523–34. doi: 10.1016/j.jacc.2020.01.052

39. Wegner FK, Plagwitz L, Doldi F, Ellermann C, Willy K, Wolfes J, et al. Machine learning in the detection and management of atrial fibrillation. *Clin Res Cardiol*. (2022) 111(9):1010–7. doi: 10.1007/s00392-022-02012-3

40. Kotb A, Armstrong S, Koev I, Antoun I, Vali Z, Panchal G, et al. Digitally enabled acute care for atrial fibrillation: conception, feasibility and early outcomes of an AF virtual ward. *Open Heart*. (2023) 10(1):e002272. doi: 10.1136/openhrt-2023-002272

41. Chan NY. Systematic screening for atrial fibrillation in the community: evidence and obstacles. *Arrhythm Electrophysiol Rev*. (2018) 7(1):39–42. doi: 10.15420/aer.2017.47.2

42. Kirchhof P, Blank BF, Calvert M, Camm AJ, Chlouverakis G, Diener H-C, et al. Probing oral anticoagulation in patients with atrial high rate episodes: rationale and design of the non-vitamin K antagonist oral anticoagulants in patients with atrial high rate episodes (NOAH-AFNET 6) trial. *Am Heart J*. (2017) 190:12–8. doi: 10.1016/j.ahj.2017.04.015

43. McIntyre WF, Benz AP, Becher N, Healey JS, Granger CB, Rivard L, et al. Direct oral anticoagulants for stroke prevention in patients with device-detected atrial fibrillation: a study-level meta-analysis of the NOAH-AFNET 6 and ARTESiA trials. *Circulation*. (2024) 149(13):981–8. doi: 10.1161/CIRCULATIONAHA.123.067512

44. Xu Y, Zhang S, Cao Z, Chen Q, Xiao W. Extreme learning machine for heartbeat classification with hybrid time-domain and wavelet time-frequency features. *J Healthc Eng*. (2021) 2021(1):6674695. doi: 10.1155/2021/6674695

45. Singh R, Rajpal N, Mehta R. An empiric analysis of wavelet-based feature extraction on deep learning and machine learning algorithms for arrhythmia classification. *Int J Interact Multimed Artif Intell*. (2021) 6:25–34. doi: 10.9781/ijimai.2020.11.005

46. Abdullh LA, Al-Ani MS. CNN-LSTM based model for ECG arrhythmias and myocardial infarction classification. *Adv Sci Technol Eng Syst*. (2020) 5(5):601–6. doi: 10.25046/aj050573

47. Ullah A, Anwar SM, Bilal M, Mehmood RM. Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation. *Remote Sens (Basel)*. (2020) 12(10):1685. doi: 10.3390/rs12101685

48. Zhao Y, Cheng J, Zhang P, Peng X. ECG Classification using deep CNN improved by wavelet transform. *Comput Mater Con.* (2020) 64:1615–28. doi: 10.32604/cmc.2020.09938
49. Nurmaini S, Tondas AE, Darmawahyuni A, Rachmatullah MN, Partan RU, Firdaus F, et al. Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. *Future Gener Comput Syst.* (2020) 113:304–17. doi: 10.1016/j.future.2020.07.021
50. Kora P, Kumari CU, Meenakshi K. Atrial fibrillation detection using discrete wavelet transform. *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICEECT)* (2019) IEEE.
51. Chashmi AJ, Amirani MC. An efficient and automatic ECG arrhythmia diagnosis system using DWT and HOS features and entropy-based feature selection procedure. *J Electr Bioimp.* (2019) 10(1):47. doi: 10.2478/joeb-2019-0007
52. Alarsan FI, Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *J Big Data.* (2019) 6(1):1–15. doi: 10.1186/s40537-019-0244-x
53. Anwar SM, Gul M, Majid M, Alnowami M. Arrhythmia classification of ECG signals using hybrid features. *Comput Math Methods Med.* (2018) 2018(1):1380348. doi: 10.1155/2018/1380348
54. Lassoued H, Ketata R. Artificial neural network classifier for heartbeat arrhythmia detection. *Proc Eng Technol.* (2017) 22:67–72.
55. Xin Y, Zhao Y. Paroxysmal atrial fibrillation recognition based on multi-scale wavelet α -entropy. *Biomed Eng Online.* (2017) 16:1–12. doi: 10.1186/s12938-017-0406-z
56. Filos D, Chouvarda I, Tachmatzidis D, Vassilikos V, Maglaveras N. Beat-to-beat P-wave morphology as a predictor of paroxysmal atrial fibrillation. *Comput Methods Programs Biomed.* (2017) 151:111–21. doi: 10.1016/j.cmpb.2017.08.016
57. Kaya Y, Pehlivan H, Tenekeci M. Effective ECG beat classification using higher order statistic features and genetic feature selection. *Biomed Res India.* (2017) 28(17):7594–603.
58. Saraswat S, Srivastava G, Shukla SN. Wavelet transform based feature extraction and classification of atrial fibrillation arrhythmia. *Biomed Pharmacol J.* (2017) 10(4):1715–25. doi: 10.13005/bpj/1284
59. Dewangan NK, Shukla S. ECG Arrhythmia classification using discrete wavelet transform and artificial neural network. *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (2016) IEEE.
60. Thomas M, Das MK, Ari S. Automatic ECG arrhythmia classification using dual tree complex wavelet based features. *AEU Int J Electron Commun.* (2015) 69(4):715–21. doi: 10.1016/j.aue.2014.12.013
61. Barmase S, Das S, Mukhopadhyay S. Wavelet transform-based analysis of QRS complex in ECG signals. *arXiv preprint arXiv.* (2013):13116460.
62. Sarkaleh MK, Shahbahrami A. Classification of ECG arrhythmias using discrete wavelet transform and neural networks. *Int J Comput Sci Eng Appl.* (2012) 2(1):1. doi: 10.5121/ijcsea.2012.2101
63. Kim J, Min SD, Lee M. An arrhythmia classification algorithm using a dedicated wavelet adapted to different subjects. *Biomed Eng Online.* (2011) 10:1–19. doi: 10.1186/1475-925X-10-56
64. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med.* (2019) 381(2):1909–17. doi: 10.1056/NEJMoa1901183
65. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. *JAMA.* (2018) 320(2):146–55. doi: 10.1001/jama.2018.8102
66. Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C, et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. *Circulation.* (2017) 136(19):1784–94. doi: 10.1161/CIRCULATIONAHA.117.030583
67. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, Quinn FR, Hummers E, Ivers N, et al. Screening for atrial fibrillation in the older population: a randomized clinical trial. *JAMA Cardiol.* (2021) 6(5):558–67. doi: 10.1001/jamacardio.2021.0038
68. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit heart study. *Circulation.* (2022) 146(19):1415–24. doi: 10.1161/CIRCULATIONAHA.122.060291
69. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. *J Am Coll Cardiol.* (2019) 74(19):2365–75. doi: 10.1016/j.jacc.2019.08.019
70. Lip GY, Nieuwlaet R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* (2010) 137(2):263–72. doi: 10.1378/chest.09-1584
71. Ru X, Wang T, Zhu L, Ma Y, Qian L, Sun H, et al. Using a clinical decision support system to improve anticoagulation in patients with nonvalve atrial fibrillation in China's primary care settings: a feasibility study. *Int J Clin Pract.* (2023) 2023(1):2136922. doi: 10.1155/2023/2136922
72. Han L, Askari M, Altman RB, Schmitt SK, Fan J, Bentley JP, et al. Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. *Circ Cardiovasc Qual Outcomes.* (2019) 12(10):e005595. doi: 10.1161/CIRCOUTCOMES.118.005595
73. Popat A, Yadav S, Obholz J, Hwang EA, Rehman AU, Sharma P. The efficacy of artificial intelligence in the detection and management of atrial fibrillation. *Cureus.* (2025) 17(1):e77135. doi: 10.7759/cureus.77135
74. Hiraoka D, Inui T, Kawakami E, Oya M, Tsuji A, Honma K, et al. Diagnosis of atrial fibrillation using machine learning with wearable devices after cardiac surgery: algorithm development study. *JMIR Form Res.* (2022) 6(8):e35396. doi: 10.2196/35396
75. Lin Q, Li T, Shakeel PM, Samuel RDJ. Advanced artificial intelligence in heart rate and blood pressure monitoring for stress management. *J Ambient Intell Humaniz Comput.* (2021) 12:3329–40. doi: 10.1007/s12652-020-02650-3
76. Sanchez de la Nava AM, Atienza F, Bermejo J, Fernandez-Aviles F. Artificial intelligence for a personalized diagnosis and treatment of atrial fibrillation. *Am J Physiol Heart Circ Physiol.* (2021) 320(4):H1337–47. doi: 10.1152/ajpheart.00764.2020
77. Lassau N, Bousaid I, Chouzenoux E, Lamarque J-P, Charmettant B, Azoulay M, et al. Three artificial intelligence data challenges based on CT and MRI. *Diagn Interv Imaging.* (2020) 101(12):783–8. doi: 10.1016/j.diii.2020.03.006
78. Alonso A, Krijthe BP, Aspelund T, Stepan KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc.* (2013) 2(2):e000102. doi: 10.1161/JAHA.112.000102
79. Santala OE, Halonen J, Martikainen S, Jäntti H, Rissanen TT, Tarvainen MP, et al. Automatic mobile health arrhythmia monitoring for the detection of atrial fibrillation: prospective feasibility, accuracy, and user experience study. *JMIR Mhealth Uhealth.* (2021) 9(10):e29933. doi: 10.2196/29933
80. Zimmerman RM, Hernandez EJ, Watkins WS, Blue N, Tristani-Firouzi M, Yandell M, et al. An explainable artificial intelligence approach for discovering social determinants of health and risk interactions for stroke in patients with atrial fibrillation. *Am J Cardiol.* (2023) 201:224–6. doi: 10.1016/j.amjcard.2023.05.064
81. Ngufor C, Zhang N, Houten HKV, Holmes DR, Graff-Radford J, Alkhouli M, et al. Causal machine learning for left atrial appendage occlusion in patients with atrial fibrillation. *JACC Clin Electrophysiol.* (2025) 11(5):977–86. doi: 10.1016/j.jacep.2024.12.013
82. Levy AE, Biswas M, Weber R, Tarakji K, Chung M, Noseworthy PA, et al. Applications of machine learning in decision analysis for dose management for dofetilide. *PLoS One.* (2019) 14(12):e0227324. doi: 10.1371/journal.pone.0227324
83. Mehta N, Devarakonda MV. Machine learning, natural language programming, and electronic health records: the next step in the artificial intelligence journey? *J Allergy Clin Immunol.* (2018) 141(6):2019–21.e1. doi: 10.1016/j.jaci.2018.02.025
84. Winkle RA, Mead RH, Engel G, Patrawala RA. Long-term results of atrial fibrillation ablation: the importance of all initial ablation failures undergoing a repeat ablation. *Am Heart J.* (2011) 162(1):193–200. doi: 10.1016/j.ahj.2011.04.013
85. Liu C-M, Chang S-L, Chen H-H, Chen W-S, Lin Y-J, Lo L-W, et al. The clinical application of the deep learning technique for predicting trigger origins in patients with paroxysmal atrial fibrillation with catheter ablation. *Circ Arrhythm Electrophysiol.* (2020) 13(11):e008518. doi: 10.1161/CIRCEP.120.008518
86. Müller J, Nentwich K, Berkovitz A, Ene E, Sonne K, Zhuravlev V, et al. Acute oesophageal safety and long-term follow-up of AI-guided high-power short-duration with 50 W for atrial fibrillation ablation. *Europace.* (2023) 25(4):1379–91. doi: 10.1093/europace/euad053
87. Sanchez de la Nava AM, Arenal Á, Fernández-Avilés F, Atienza F. Artificial intelligence-driven algorithm for drug effect prediction on atrial fibrillation: an in silico population of models approach. *Front Physiol.* (2021) 12:768468. doi: 10.3389/fphys.2021.768468
88. Deisenhofer I, Albenque J-P, Busch S, Gitenay E, Mountantonakis S, Roux A, et al. LB-469805-01 tailored cardiac ablation procedure for persistent atrial fibrillation guided by artificial intelligence: the tailored-AF randomized clinical trial. *Heart Rhythm.* (2024) 21(7):1199. doi: 10.1016/j.hrthm.2024.04.025
89. Seitz J, Durdez TM, Albenque JP, Pisapia A, Gitenay E, Durand C, et al. Artificial intelligence software standardizes electrogram-based ablation outcome for persistent atrial fibrillation. *J Cardiovasc Electrophysiol.* (2022) 33(11):2250–60. doi: 10.1111/jce.15657
90. Bahlke F, Englert F, Popa M, Bourrier F, Reents T, Lennerz C, et al. First clinical data on artificial intelligence-guided catheter ablation in long-standing persistent atrial fibrillation. *J Cardiovasc Electrophysiol.* (2024) 35(3):406–14. doi: 10.1111/jce.16184
91. Antoun I, Layton GR, Nizam A, Barker J, Abdelrazik A, Eldesouky M, et al. Hypertension and atrial fibrillation: bridging the gap between mechanisms, risk, and therapy. *Medicina (Kaunas).* (2025) 61(2):362. doi: 10.3390/medicina61020362

92. Serhal H, Abdallah N, Marion J-M, Chauvet P, Oueidat M, Humeau-Heurtier A. Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG. *Comput Biol Med.* (2022) 142:105168. doi: 10.1016/j.compbio.2021.105168
93. Boriani G, Mei DA, Lip GYH, Consortium TA. Artificial intelligence in patients with atrial fibrillation to manage clinical complexity and comorbidities: the ARISTOTELES project. *Eur Heart J.* (2024) 46(9):775–7. doi: 10.1093/eurheartj/ehae792
94. Kim RS, Simon S, Powers B, Sandhu A, Sanchez J, Borne RT, et al. Machine learning methodologies for prediction of rhythm-control strategy in patients diagnosed with atrial fibrillation: observational, retrospective, case-control study. *JMIR Med Inform.* (2021) 9(12):e29225. doi: 10.2196/29225
95. Deisenhofer I, Albenque J-P, Busch S, Gitenay E, Mountantonakis SE, Roux A, et al. Artificial intelligence for individualized treatment of persistent atrial fibrillation: a randomized controlled trial. *Nat Med.* (2025) 31:1–8. doi: 10.1038/s41591-025-03517-w
96. Grégoire J-M, Gilon C, Marelli F, Bersini H, Carlier S. Genomics vs. AI-enhanced electrocardiogram: predicting atrial fibrillation in the era of precision medicine. *Explor Digit Health Technol.* (2025) 3:101141. doi: 10.37349/edht.2025.101141
97. Isaksen JL, Baumert M, Hermans AN, Maleckar M, Linz D. Artificial intelligence for the detection, prediction, and management of atrial fibrillation. *Herzschrittmacherther Elektrophysiol.* (2022) 33(1):34–41. doi: 10.1007/s00399-022-00839-x
98. Matias I, Garcia N, Pirbhulal S, Felizardo V, Pombo N, Zacarias H, et al. Prediction of atrial fibrillation using artificial intelligence on electrocardiograms: a systematic review. *Comput Sci Rev.* (2021) 39:100334. doi: 10.1016/j.cosrev.2020.100334
99. Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. *Wiley Interdiscip Rev Data Min Knowl Discov.* (2021) 11(5):e1424. doi: 10.1002/widm.1424
100. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev.* (2022) 55:1–66. doi: 10.1007/s10462-021-10088-y
101. Prasitlumkum N, Cheungpasitporn W, Chokesuwattanaskul A, Thangjui S, Thongprayoon C, Bathini T, et al. Diagnostic accuracy of smart gadgets/wearable devices in detecting atrial fibrillation: a systematic review and meta-analysis. *Arch Cardiovasc Dis.* (2021) 114(1):4–16. doi: 10.1016/j.acvd.2020.05.015
102. Yan BP, Lai WHS, Chan CKY, Chan SCH, Chan LH, Lam KM, et al. Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals. *J Am Heart Assoc.* (2018) 7(8):e008585. doi: 10.1161/JAHA.118.008585
103. Couderc JP, Page A, Lutz M, Tsouri GR, Hall B. Assessment of facial video-based detection of atrial fibrillation across human complexion. *Cardiovasc Digit Health J.* (2022) 3(6):305–12. doi: 10.1016/j.cvdhj.2022.08.003
104. Guhl E, Althouse AD, Pusateri AM, Kimani E, Paasche-Orlow MK, Bickmore TW, et al. The atrial fibrillation health literacy information technology trial: pilot trial of a mobile health app for atrial fibrillation. *JMIR cardio.* (2020) 4(1):e17162. doi: 10.2196/17162
105. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol.* (2012) 46(4):268–77. doi: 10.2345/0899-8205-46.4.268
106. Babel A, Taneja R, Mondello Malvestiti F, Monaco A, Donde S. Artificial intelligence solutions to increase medication adherence in patients with non-communicable diseases. *Front Digit Health.* (2021) 3:669869. doi: 10.3389/fdgh.2021.669869
107. Taniguchi H, Takata T, Takechi M, Furukawa A, Iwasawa J, Kawamura A, et al. Explainable artificial intelligence model for diagnosis of atrial fibrillation using Holter electrocardiogram waveforms. *Int Heart J.* (2021) 62(3):534–9. doi: 10.1536/ihj.21-094
108. Jo Y-Y, Cho Y, Lee SY, Kwon J-m, Kim K-H, Jeon K-H, et al. Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *Int J Cardiol.* (2021) 328:104–10. doi: 10.1016/j.ijcard.2020.11.053
109. Ong JCL, Seng BJJ, Law JZF, Low LL, Kwa ALH, Giacomini KM, et al. Artificial intelligence, ChatGPT, and other large language models for social determinants of health: current state and future directions. *Cell Rep Med.* (2024) 5(1):101356. doi: 10.1016/j.xcrm.2023.101356
110. Chen RJ, Wang JJ, Williamson DF, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng.* (2023) 7(6):719–42. doi: 10.1038/s41551-023-01056-8
111. Jones NR, Taylor CJ, Hobbs FR, Bowman L, Casadei B. Screening for atrial fibrillation: a call for evidence. *Eur Heart J.* (2020) 41(10):1075–85. doi: 10.1093/eurheartj/ehz834
112. Kumar D, Puthusserypady S, Dominguez H, Sharma K, Bardram JE. An investigation of the contextual distribution of false positives in a deep learning-based atrial fibrillation detection algorithm. *Expert Syst Appl.* (2023) 211:118540. doi: 10.1016/j.eswa.2022.118540
113. Safdar NM, Banja JD, Meltzer CC. Ethical considerations in artificial intelligence. *Eur J Radiol.* (2020) 122:108768. doi: 10.1016/j.ejrad.2019.108768
114. Bostrom N. Ethical issues in advanced artificial intelligence. In: Wallach W, Asaro P, editors. *Machine Ethics and Robot Ethics*. 1st edn. London: Routledge. (2020). p. 69–75. doi: 10.4324/9781003074991-7
115. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform.* (2019) 28(01):128–34. doi: 10.1055/s-0039-1677903
116. Park SH, Choi J, Byeon J-S. Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence. *Korean J Radiol.* (2021) 22(3):442. doi: 10.3348/kjr.2021.0048
117. Jiao Z, Choi JW, Halsey K, Tran TML, Hsieh B, Wang D, et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. *Lancet Digit Health.* (2021) 3(5):e286–94. doi: 10.1016/S2589-7500(21)00039-X
118. Zhao Y, Li Y, Xing L, Lei H, Chen D, Tang C, et al. The performance of artificial intelligence in cervical colposcopy: a retrospective data analysis. *J Oncol.* (2022) 2022(1):4370851. doi: 10.1155/2022/4370851
119. Ruskin KJ, Ruskin AC. Improving alarms and alerts for the next generation of medical devices. *Anesth Analg.* (2022). doi: 10.1213/ANE.00000000000007420
120. Hedderich DM, Keicher M, Wiestler B, Gruber MJ, Burwinkel H, Hinterwimmer F, et al. AI for doctors—a course to educate medical professionals in artificial intelligence for medical imaging. *Healthcare (Basel).* (2021) 9:1278. doi: 10.3390/healthcare9101278
121. Noseworthy PA, Attia ZI, Behnken EM, Giblon RE, Bews KA, Liu S, et al. Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. *Lancet.* (2022) 400(10359):1206–12. doi: 10.1016/S0140-6736(22)01637-3
122. Nemati S, Ghassemi MM, Ambai V, Isakadze N, Levantsevych O, Shah A, et al. Monitoring and detecting atrial fibrillation using wearable technology. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2016) IEEE.



OPEN ACCESS

EDITED BY

Rizwan Qureshi,
Hamad Bin Khalifa University, Qatar

REVIEWED BY

Li-Da Wu,
Nanjing Medical University, China
Carlo Lavallo,
Sapienza University of Rome, Italy

*CORRESPONDENCE

Panteleimon Pantelidis
✉ ppantele@med.uoa.gr

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 04 July 2025

ACCEPTED 30 September 2025

PUBLISHED 16 October 2025

CITATION

Pantelidis P, Ruipérez-Campillo S, Vogt JE, Antonopoulos A, Gialamas I, Zakynthinos GE, Spartalis M, Dilaveris P, Millet J, Papaioannou TG, Oikonomou E and Siasos G (2025) ECG-XPLAIM: eXplainable Locally-adaptive Artificial Intelligence Model for arrhythmia detection from large-scale electrocardiogram data. *Front. Cardiovasc. Med.* 12:1659971. doi: 10.3389/fcvm.2025.1659971

COPYRIGHT

© 2025 Pantelidis, Ruipérez-Campillo, Vogt, Antonopoulos, Gialamas, Zakynthinos, Spartalis, Dilaveris, Millet, Papaioannou, Oikonomou and Siasos. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ECG-XPLAIM: eXplainable Locally-adaptive Artificial Intelligence Model for arrhythmia detection from large-scale electrocardiogram data

Panteleimon Pantelidis^{1*†}, Samuel Ruipérez-Campillo^{2†}, Julia E. Vogt², Alexios Antonopoulos³, Ioannis Gialamas¹, George E. Zakynthinos¹, Michael Spartalis¹, Polychronis Dilaveris^{1,3}, Jose Millet⁴, Theodore G. Papaioannou³, Evangelos Oikonomou¹ and Gerasimos Siasos¹

¹3rd Department of Cardiology, National and Kapodistrian University of Athens, Athens, Greece,

²Department of Computer Science, ETH Zurich, Zurich, Switzerland, ³1st Department of Cardiology, National and Kapodistrian University of Athens, Athens, Greece, ⁴ITACA Institute, Universitat Politècnica de Valencia, Valencia, Spain

Background: Timely and accurate detection of arrhythmias from electrocardiograms (ECGs) is crucial for improving patient outcomes. While artificial intelligence (AI)-based ECG classification has shown promising results, limited transparency and interpretability often impede clinical adoption.

Methods: We present ECG-XPLAIM, a novel deep learning model dedicated to ECG classification that employs a one-dimensional inception-style convolutional architecture to capture local waveform features (e.g., waves and intervals) and global rhythm patterns. To enhance interpretability, we integrate Grad-CAM visualization, highlighting key waveform segments that drive the model's predictions. ECG-XPLAIM was trained on the MIMIC-IV dataset and externally validated on PTB-XL for multiple arrhythmias, including atrial fibrillation (AFib), sinus tachycardia (STach), conduction disturbances (RBBB, LBBB, LAFB), long QT (LQT), Wolff-Parkinson-White (WPW) pattern, and paced rhythm detection. We evaluated performance using sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC), and benchmarked against a simplified convolutional neural network, a two-layer gated recurrent unit (GRU), and an external, pre-trained, ResNet-based model.

Results: Internally (MIMIC-IV), ECG-XPLAIM achieved high diagnostic performance (sensitivity, specificity, AUROC > 0.9) across most tasks. External evaluation (PTB-XL) confirmed generalizability, with metric values exceeding 0.95 for AFib and STach. For conduction disturbances, macro-averaged sensitivity reached 0.90, specificity 0.95, and AUROC 0.98. Performance for LQT, WPW, and pacing rhythm detection was 0.691/0.864/0.878, 0.773/0.973/0.895, and 0.96/0.988/0.993 (sensitivity/specificity/AUROC), respectively. Compared to baseline models, ECG-XPLAIM offered superior performance across most tests, and improved sensitivity over the external ResNet-based model, albeit at the cost of specificity. Grad-CAM revealed physiologically relevant ECG segments influencing predictions and highlighted patterns of potential misclassification.

Conclusion: ECG-XPLAIM combines high diagnostic performance with interpretability, addressing a key limitation in AI-driven ECG analysis. The open-source release of ECG-XPLAIM's architecture and pre-trained weights encourages broader adoption, external validation, and further refinement for diverse clinical applications.

KEYWORDS

arrhythmia, electrocardiogram, artificial intelligence, deep learning, machine learning, cardiac signals, explainability

1 Introduction

Accurate arrhythmia detection from electrocardiogram (ECG) recordings is crucial for early intervention, particularly for life-threatening conditions such as atrial fibrillation, conduction disturbances, and other arrhythmic syndromes. If left undiagnosed or untreated, these conditions can lead to severe complications that adversely affect patient morbidity and mortality. For instance, atrial fibrillation increases the risk of stroke fivefold (1), while Wolff-Parkinson-White syndrome, long QT syndrome, and other arrhythmic disorders markedly raise susceptibility to fatal cardiac events (2). Despite the importance of prompt diagnosis, ECG interpretation is highly specialized; underdiagnosis rates exceeding 50% are reported among non-cardiologists, varying with disease, population, and clinical setting (3). This highlights the pressing need for automated diagnostic tools that can support clinicians, reduce human error, and improve arrhythmia detection rates.

Artificial intelligence (AI) and deep learning (DL) have emerged as powerful tools capable of automating ECG analysis, achieving high accuracy even in detecting atypical cases. These models can recognize complex patterns in ECG signals, often outperforming rule-based algorithms and, in some scenarios, surpassing human expertise (4–7). Nonetheless, significant challenges remain. Existing DL models frequently overlook the unique attributes of ECG signals—which exhibit both repetitive waveforms and global rhythmic patterns—requiring both local (wave- and interval-level) and global (rhythm-level) analysis (8, 9). Additionally, many proposed methods are closed-source, limiting adaptation to specialized clinical applications and generalization across diverse populations (10, 11). Another critical concern is the interpretability of AI-driven decisions; clinicians must understand model reasoning before relying on these tools for patient care (12–14). Explainable AI (XAI) techniques, including Grad-CAM (15, 16), highlight the salient ECG waveform features that contribute most to the model's decisions, thus improving interpretability. However, relatively few ECG-focused architectures combine robust performance with built-in explainability.

In this study, we introduce ECG-XPLAIM—an eXplainable, Locally-adaptive Artificial Intelligence Model designed for ECG classification. By leveraging a deep inception-style convolutional architecture (17), ECG-XPLAIM is well-suited for time-series data analysis where capturing temporal dependencies is essential. The model further integrates XAI principles (13), aligning its decision-making process with the clinical reasoning that underpins ECG

interpretation. Trained on large-scale ECG datasets to ensure scalability and broad clinical applicability, ECG-XPLAIM is also released as open-source, complete with pre-trained weights. The open framework facilitates external validation, transfer learning, and customization for specialized tasks. Overall, our work aims to advance interpretable machine learning in healthcare, offering a framework that balances high diagnostic performance with transparency and adaptability, thereby enhancing trust in AI-driven ECG diagnostics.

2 Materials and methods

2.1 Study population and data sources

We employed two large-scale, publicly available ECG datasets for model development and evaluation: MIMIC-IV and PTB-XL (18–20). MIMIC-IV contains over 800,000 12-lead ECG recordings, while PTB-XL comprises more than 25,000 12-lead ECG records, each sampled at 500 Hz for a uniform duration of 10 s (Figure 1). The raw waveform data were used without additional preprocessing, aside from replacing missing values with zeros. To evaluate the impact of conventional ECG preprocessing, additional experiments applied a 0.5–40 Hz Butterworth bandpass filter and a 50 Hz notch filter prior to training and inference, while keeping all other parameters identical (see Supplementary Section S4). Demographic and patient-specific metadata were deliberately excluded from the training process to focus exclusively on ECG waveforms and to minimize bias related to patient characteristics.

2.2 Diagnostic labels and outcome selection

We defined five classification tasks, each of them corresponding to a clinically significant arrhythmic category (Table 1). The first task (tachycardia—TACHY) involved distinguishing atrial fibrillation (AFib) from sinus tachycardia (STach). The second task focused on detecting conduction disturbances (CD), encompassing right bundle branch block (RBBB), left bundle branch block (LBBB), and left anterior fascicular block (LAFB). The remaining tasks targeted the identification of long QT (LQT), the detection of Wolff-Parkinson-White (WPW) pattern, and the recognition of paced (PACE) rhythms.

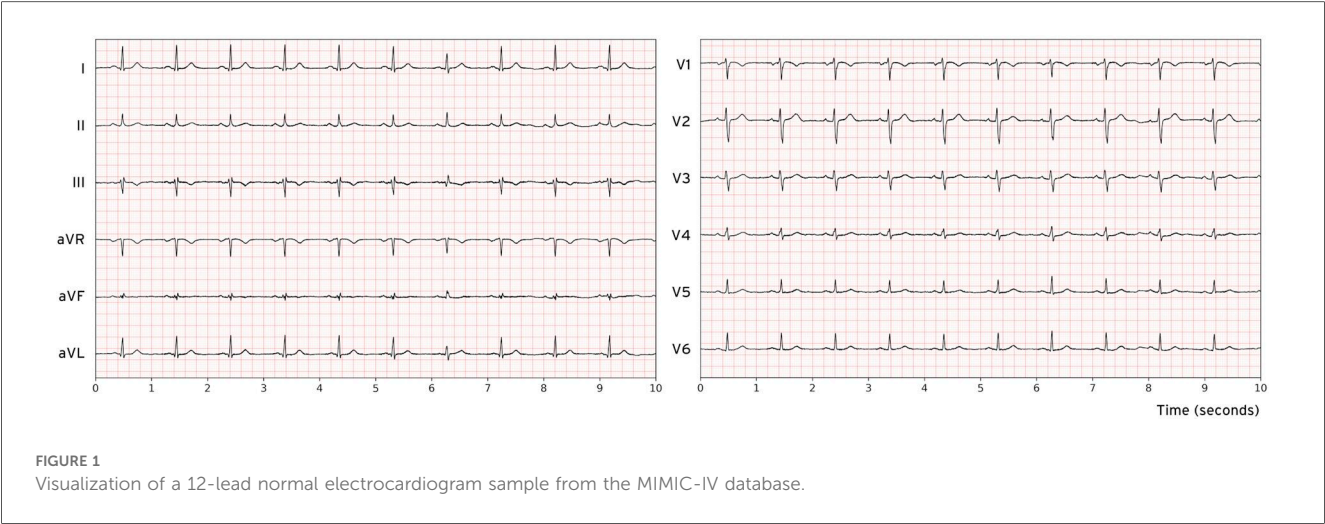


TABLE 1 Arrhythmia classification tasks and definitions.

Task	Diagnostic labels	Details
TACHY (Tachyarrhythmias)	AFib vs. STach vs. Neg ^a	Differentiates AFib, characterized by irregular RR intervals and absent P-waves, from STach, which exhibits regular RR intervals with elevated heart rates. The Neg class includes ECGs that do not show these tachyarrhythmias.
CD (Conduction Disturbances)	RBBB vs. LBBB vs. LAFB vs. Neg ^a	Detects bundle branch blocks based on QRS complex morphology and duration. RBBB is marked by a prolonged QRS (>120 ms) with an rSR' pattern in V1, LBBB by a broad QRS with deep S waves in V1, and LAFB by left-axis deviation and qR pattern in aVL. The Neg group excludes these conditions. ^b
LQT (Long QT)	LQT vs. Neg ^a	Identifies prolonged QT intervals, measured using standard correction formulas.
WPW (Wolff-Parkinson-White) pattern	WPW pattern vs. Neg ^a	Detects pre-excitation patterns characterized by short PR intervals, delta waves, and wide QRS complexes, which indicate an accessory conduction pathway.
PACE (Paced rhythm)	Paced rhythm vs. Neg ^a	Identifies cardiac pacing (atrial and/or ventricular), marked by pacemaker spikes preceding P-waves and/or QRS complexes, respectively.

AFib, atrial fibrillation; LAFB, left anterior fascicular block; LBBB, left bundle branch block; LQT, long QT; RBBB, right bundle branch block; STach, sinus tachycardia; WPW, Wolff-Parkinson-White pattern.
^aNeg, the negative class within each task does not necessarily represent normal ECGs—it may include other non-target arrhythmias or abnormalities.
^bLeft posterior fascicular block (LPFB) was excluded due to limited representation in the dataset.

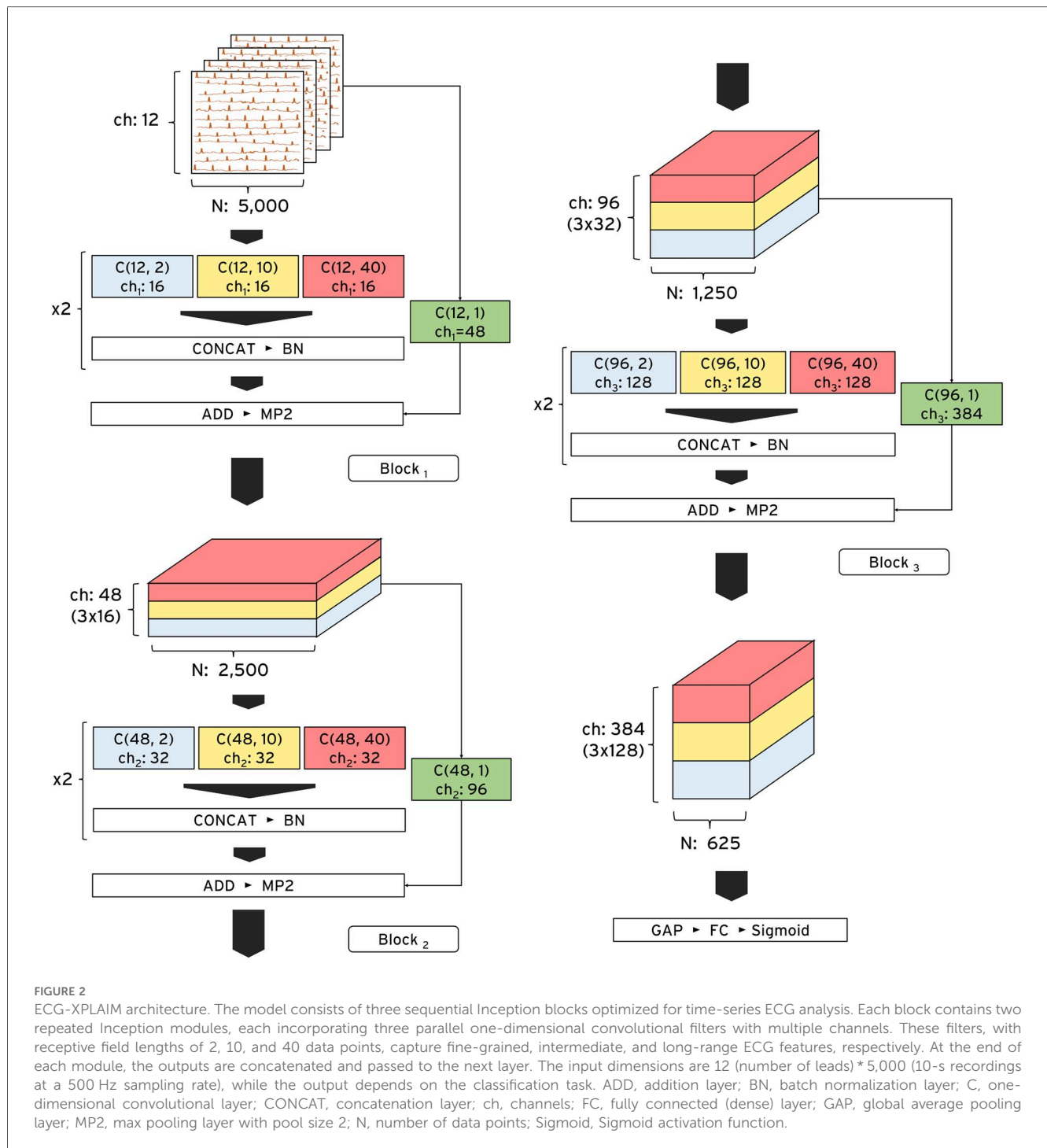
Labels were derived from structured diagnostic reports and automated ECG annotations provided within the source databases (a detailed list of key terms is presented in the [Supplementary Section S1](#)). Negative samples for each task were defined by the absence of the corresponding arrhythmic condition but could include other pathologies unrelated to the primary label. Overlapping conditions (for example, AFib coexisting with LQT or WPW overlapping with LBBB) were not excluded to reflect the complexity of real-world ECG interpretations. A manual review of the entire PTB-XL test subset, and an additional random 10% of the MIMIC-IV dataset was performed by expert clinicians to validate labeling accuracy. Agreement with database-provided labels was quantified using Cohen’s kappa with bootstrapped CIs ([Supplementary Section S5](#)).

2.3 Model design and explainability mechanism

We developed ECG-XPLAIM using a custom Inception-style convolutional neural network (CNN) architecture,

designed for time-series analysis (17, 21). This framework captures both local waveform features (e.g., P, QRS, and T waves and intervals) and global rhythm patterns, such as irregularities that underlie arrhythmias. The model comprises three residual blocks, each containing two one-dimensional Inception modules. Within each module, three convolutional kernels of lengths 2, 10, and 40 data points (corresponding to 4, 20, and 80 ms at 500 Hz sampling frequency) operate in parallel. The outputs of these parallel paths are concatenated before being fed into the subsequent layer. By incorporating skip connections, we preserved signal integrity and avoided feature degradation as the network depth increased ([Figure 2](#)). The receptive field expands progressively, roughly doubling after each block, which allows ECG-XPLAIM to detect both brief waveform and broader rhythm disturbances. This emphasizes high sensitivity for transient lesions while maintaining robust generalization across diverse ECG patterns.

To enhance interpretability, we integrated a customized one-dimensional Gradient-weighted Class Activation Mapping (Grad-CAM) mechanism tailored for multi-lead



ECGs (15). This approach generates activation heatmaps that highlight the waveform regions most influential for the model's predictions. By aggregating lead-specific Grad-CAM maps across all 12 leads, ECG-XPLAIM provides a comprehensive view of its decision-making. We overlaid these heatmaps on the raw ECG traces to underline clinically relevant features such as the presence or absence of P-wave and their morphology, QRS complex widening, and RR interval variations.

2.4 Training and evaluation strategies

We trained ECG-XPLAIM on the MIMIC-IV dataset and then evaluated internally on independent cohorts from the same dataset, while also testing externally on PTB-XL. To address class imbalance, the maximum number of samples per class was capped at 50,000, except for WPW detection, where data scarcity necessitated a 1:2 positive-to-negative ratio (600 vs. 1,220 samples). For AFib vs. STach vs. Negative, 50,000 samples

were used per class. For RBBB, LBBB, LAFB, and Negative, 25,000 were assigned to each category. For LQT vs. Negative, 39,000 samples per class were used, and for PACE vs. Negative, 30,000 per class were retained. Distinct training, validation, and test subsets were prepared to avoid overlapping. As an illustrative experiment, we further fine-tuned the pre-trained MIMIC-IV model on varying fractions of PTB-XL (0%, 5%, 10%, 20%, 50%) and evaluated it on the remaining data ([Supplementary Section S6](#)). This setup is included to demonstrate feature transferability, while it is not part of our primary claims, since it does not fulfill the requirement for independence of the external dataset ([22](#)). Training was performed in mini-batches using TensorFlow/Keras on an NVIDIA L4 GPU (24GB VRAM) for 100 epochs without early stopping. We adopted the Adaptive Moment Estimation (ADAM) optimizer ([23](#)), with an initial learning rate of 0.01, subject to exponential decay (0.95 per epoch). The batch size was 128 for most tasks but reduced to 32 for WPW detection due to fewer positive samples. We selected the optimal checkpoint based on validation metrics during training to minimize overfitting.

2.5 Diagnostic performance and statistical analysis

We evaluated model performance using sensitivity (recall), specificity, and area under the receiver operating characteristic curve (AUROC), estimating 95% confidence intervals (CIs) for each metric. Internal assessment took place on the MIMIC-IV test set, whereas external evaluation was performed on the PTB-XL cohort, ensuring no overlap with training data. CIs for sensitivity and specificity were computed using the Clopper-Pearson exact method ([24](#)), and AUROC CIs were derived via DeLong's method ([25](#)). We also evaluated task-wise operating points suitable for screening (maximizing sensitivity) and diagnostic confirmation (maximizing specificity), by scanning model decision thresholds from 0.0 to 1.0, at 0.1 increments, and reporting macro-averaged sensitivity/specificity pairs.

To establish benchmark comparisons, we trained two deep learning baselines under identical conditions. First, we implemented a conventional one-dimensional CNN model (vanilla CNN) with three standard convolutional layers followed by batch normalization and max pooling. Second, we applied a more advanced, double-layered, gated recurrent unit (GRU) architecture, designed for time-series feature extraction (implementation details are provided in [Supplementary Section S2](#)) ([26](#)). We also compared against an external, pre-trained, ResNet-based deep network that had been previously validated for 12-lead ECG classification ([27](#)). This external model was adapted to match our sampling frequency, lead configuration, and data settings. As the commonly supported diagnostic categories were only four (AFib, STach, RBBB, and LBBB), comparisons were restricted to these conditions.

We performed pairwise statistical analyses between ECG-XPLAIM and the counterpart models, across all classification tasks, to identify significant differences in performance. We used

McNemar's test to compare model sensitivity and specificity ([28](#), [29](#)), and applied bootstrap resampling ($n=1,000$) to compare AUROC differences ([30](#)), employing a significance threshold of 0.05. To assess interpretability, we generated Grad-CAM heatmaps for correctly classified and misclassified recordings, examining which waveform components informed the model's predictions. We compared these heatmaps across all tasks, offering insight into potential biases and failure modes.

3 Results

3.1 Diagnostic performance evaluation

3.1.1 Internal evaluation

ECG-XPLAIM demonstrated robust classification performance across all arrhythmia detection tasks in the MIMIC-IV test set ([Table 2](#)). For tachyarrhythmia classification, it achieved a sensitivity of 0.897 (95% CI: 0.876–0.915) for AFib, and 0.95 (95% CI: 0.935–0.963) for STach. Specificity and AUROC values were greater or equal to 0.94 and 0.98, respectively. When detecting conduction disturbances (RBBB, LBBB, and LAFB), the model yielded macro-averaged sensitivity, specificity, and AUROC of 0.941, 0.972, and 0.993, respectively, demonstrating consistent performance across these abnormalities.

ECG-XPLAIM achieved a sensitivity of 0.93 (95% CI: 0.912–0.945) and specificity of 0.897 (95% CI: 0.876–0.915) for LQT detection, with an AUROC of 0.969 (95% CI: 0.962–0.977). For WPW pattern, it attained a sensitivity of 0.99 (95% CI: 0.946–1) and specificity of 0.95 (95% CI: 0.887–0.984), translating to an AUROC of 0.992 (95% CI: 0.980–1). PACE detection yielded a sensitivity of 0.927 (95% CI: 0.909–0.942) and specificity of 0.983 (95% CI: 0.973–0.99), with an AUROC of 0.985 (95% CI: 0.979–0.99). Overall, macro-averaged AUROC values exceeded 0.96 for all tasks in the internal evaluation, as illustrated in [Figure 3](#).

3.1.2 External evaluation

Evaluation on PTB-XL, which was not used during model development, confirmed the strong generalizability of ECG-XPLAIM for most arrhythmias ([Table 2](#)). In AFib and STach classification, sensitivity values were 0.954 (95% CI: 0.942–0.964) and 0.956 (95% CI: 0.940–0.969), respectively, with specificity of 0.964 (95% CI: 0.955–0.971) and 0.974 (95% CI: 0.967–0.979), and AUROC of 0.988 (95% CI: 0.984–0.991) and 0.991 (95% CI: 0.988–0.994). Conduction disturbance detection remained similarly robust, with RBBB identified at a sensitivity of 0.996 (95% CI: 0.986–1) and specificity of 0.966 (95% CI: 0.959–0.972). Although LBBB achieved a sensitivity of 0.99 (95% CI: 0.977–0.997), its specificity was lower at 0.927 (95% CI: 0.918–0.936). LAFB exhibited a reduced sensitivity of 0.714 (95% CI: 0.691–0.736), but maintained a high specificity of 0.962 (95% CI: 0.954–0.970).

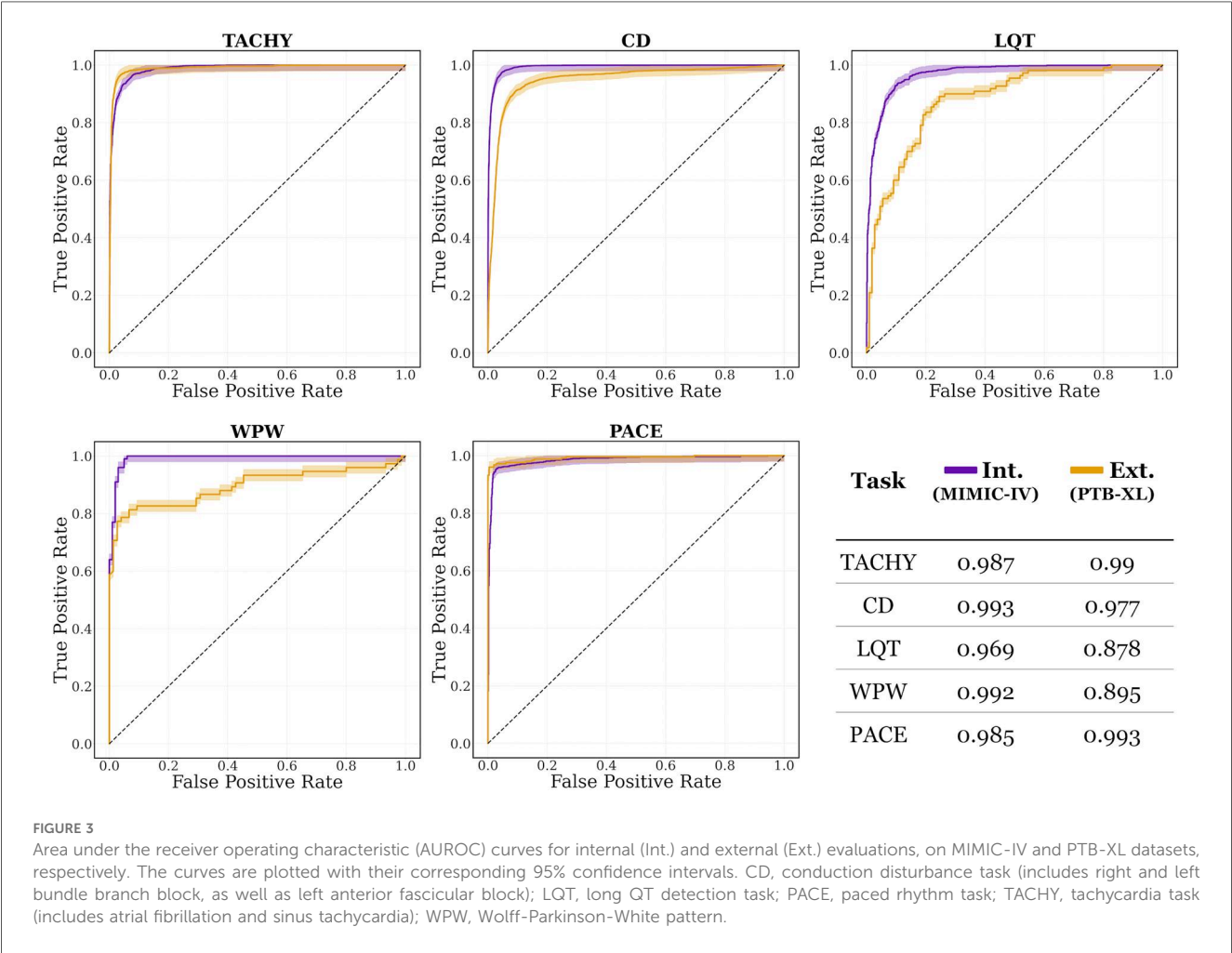
LQT detection showed a modest decline in performance compared to internal evaluation, with a sensitivity of 0.691 (95% CI: 0.596–0.776), specificity of 0.864 (95% CI: 0.785–0.922), and

TABLE 2 Diagnostic performance of ECG-XPLAIM in internal and external evaluation.

Task	Internal evaluation (MIMIC-IV)			External evaluation (PTB-XL)		
	Sensitivity	Specificity	AUROC	Sensitivity	Specificity	AUROC
TACHY						
AFib	0.897 (0.876, 0.915)	0.939 (0.928, 0.949)	0.976 (0.97, 0.981)	0.954 (0.942, 0.964)	0.964 (0.955, 0.971)	0.988 (0.984, 0.991)
STach	0.95 (0.935, 0.963)	0.978 (0.97, 0.984)	0.995 (0.992, 0.997)	0.956 (0.94, 0.969)	0.974 (0.967, 0.979)	0.991 (0.988, 0.994)
macro-avg	0.924	0.958	0.987	0.955	0.969	0.99
CD						
RBBB	0.95 (0.938, 0.961)	0.974 (0.967, 0.98)	0.994 (0.992, 0.997)	0.996 (0.986, 1)	0.966 (0.959, 0.972)	0.994 (0.992, 0.997)
LBBB	0.945 (0.929, 0.958)	0.982 (0.976, 0.986)	0.995 (0.993, 0.998)	0.99 (0.977, 0.997)	0.927 (0.918, 0.936)	0.99 (0.987, 0.993)
LAFB	0.928 (0.913, 0.941)	0.961 (0.953, 0.968)	0.989 (0.985, 0.992)	0.714 (0.691, 0.736)*	0.962 (0.954, 0.97)	0.946 (0.939, 0.953)
macro-avg	0.941	0.972	0.993	0.9	0.952	0.977
LQT	0.93 (0.912, 0.945)	0.897 (0.876, 0.915)	0.969 (0.962, 0.977)	0.691 (0.596, 0.776)*	0.864 (0.785, 0.922)	0.878 (0.835, 0.922)
WPW	0.99 (0.946, 1)	0.95 (0.887, 0.984)	0.992 (0.98, 1)	0.773 (0.662, 0.862)*	0.973 (0.907, 0.997)	0.895 (0.846, 0.944)
PACE	0.927 (0.909, 0.942)	0.983 (0.973, 0.99)	0.985 (0.979, 0.99)	0.96 (0.928, 0.981)	0.988 (0.965, 0.998)	0.993 (0.985, 1)

AFib, atrial fibrillation; AUROC, area under the receiver operating characteristic; CD, conduction disturbance task; LAFB, left anterior fascicular block; LBBB, left bundle branch block; LQT, long QT; Macro-avg, macro-averaged metric; PACE, Paced rhythm; RBBB, right bundle branch block; STach, Sinus tachycardia; TACHY, Tachycardia task; WPW, Wolff-Parkinson-White pattern.

Metrics are reported with 95% confidence intervals; The exact number of samples per class for each task is provided in the [Supplementary Material](#); Metrics marked with an asterisk (*) indicate values below 0.9 with 95% confidence.



an AUROC of 0.878 (95% CI: 0.835–0.922). Detection of WPW pattern also revealed a drop in sensitivity to 0.773 (95% CI: 0.662–0.862), with specificity remaining high at [0.973 (95% CI: 0.907–0.997)]. Paced rhythm detection achieved a sensitivity of 0.96 [95% CI: 0.928–0.981], specificity of 0.988 (95% CI: 0.965–0.998), and AUROC of 0.993 (95% CI: 0.985–1).

Overall, while certain conditions (notably LQT and WPW) showed reduced sensitivity on the external dataset, AUROC values persisted between 0.88 and 0.99 across tasks, confirming the model's strong generalization capabilities (Figure 3).

3.1.3 Supplementary analyses

We explored the effect of preprocessing (bandpass, notch), showing minimal or adverse impact on performance (Supplementary Table S3). For example, pacing detection sensitivity fell when filtering likely attenuated pacing spikes. Inter-observer agreement confirmed excellent label quality (Cohen's $\alpha \geq 0.99$ across all classes; Supplementary Table S4). Transfer learning experiments demonstrated that fine-tuning with small PTB-XL fractions could improve sensitivity for some tasks (e.g., WPW), but at the cost of specificity for some classes such as LQT, consistent with overfitting (Supplementary Table S5, Supplementary Figure S3). Threshold scans revealed balanced operating points at thresholds 0.4–0.6 for specific tasks, such as TACHY, CD and PACE, while for the rest, particularly for LQT, context-dependent trade-offs were illustrated (Supplementary Table S6).

3.1.4 Explainability analysis

We performed Grad-CAM-based analysis to visualize the waveform regions that informed ECG-XPLAIM's classification decisions. The model highlighted clinically meaningful features for each arrhythmia, while misclassifications offered insights into the model's potential attention biases, comprising focus on specific wave morphology abnormalities and interval duration deviations. Figure 4 presents single- or few-lead visualizations of selected correctly and incorrectly classified cases, highlighting the waveform regions that influenced model predictions. A more extensive demonstration, including 16 full-lead case studies with heatmaps, is available in Supplementary Section S8.

3.2 Performance benchmarking

ECG-XPLAIM outperformed both baselines across most tasks, while, compared to the pre-trained model, ECG-XPLAIM exhibited superior sensitivity but slightly lower specificity, suggesting a tendency to minimize false negatives (Table 3).

In AFib detection, ECG-XPLAIM reached an AUROC of 0.958 (95% CI: 0.951–0.965), surpassing the vanilla CNN (0.835, 95% CI: 0.823–0.848) and GRU model (0.951, 95% CI: 0.943–0.958), while achieving comparable performance to the external model (0.964, 95% CI: 0.956–0.971). Sensitivity was 0.964 (95% CI: 0.944–0.979), significantly higher or comparable against all baselines, and specificity (0.849, 95% CI: 0.835–0.861) was higher than CNN and GRU, but slightly lower than the external model. For

STach, ECG-XPLAIM achieved a sensitivity of 0.966 (95% CI: 0.946–0.980), a specificity of 0.949 (95% CI: 0.941–0.957), and an AUROC of 0.983 (95% CI: 0.979–0.988). Compared to the external model, ECG-XPLAIM reported higher sensitivity and slightly lower specificity, while AUROC was similar.

In conduction disturbances, ECG-XPLAIM maintained superior or comparable performance relative to the CNN and GRU baselines in most categories, with a slight exception in LAFB, where the GRU displayed a marginally higher AUROC. Sensitivity in detecting RBBB and LBBB was significantly higher than that of the external model (0.996, 95% CI: 0.986–1 for RBBB and 0.988, 95% CI: 0.974–0.996 for LBBB), although specificity was somewhat lower. For long QT (LQT) detection, ECG-XPLAIM achieved an AUROC of 0.81 (95% CI: 0.797–0.823), outperforming the CNN (0.727, 95% CI: 0.713–0.742) and GRU (0.675, 95% CI: 0.660–0.691). It also showed higher AUROC for WPW (0.863, 95% CI: 0.852–0.874) compared to CNN (0.562, 95% CI: 0.546–0.579) and GRU (0.521, 95% CI: 0.504–0.537), yielding also better sensitivity and specificity. For paced rhythm detection, ECG-XPLAIM surpassed both baselines in AUROC, with a value of 0.985 (95% CI: 0.981–0.989), while delivering a markedly higher specificity than both other models and comparable sensitivity. A summary of these comparative results is illustrated in Figure 5.

3.3 User experience and integration

To facilitate the adoption and practical utilization of ECG-XPLAIM, in both research and clinical environments, we provide pre-trained model weights and ready-to-use implementations for each classification task. We also provide the source code of model architecture, along with detailed documentation. A step-by-step user guide has been developed to assist clinicians and researchers in utilizing ECG-XPLAIM, outlining input formatting requirements, framework specifications, inference execution, and interpretation of outputs. Additionally, a dedicated Grad-CAM visualization module is included to support explainability assessment. This module enables users to generate heatmaps themselves that can point to ECG regions of importance, providing transparency into the model's decision-making process. By providing these tools and resources, we aim to position ECG-XPLAIM as a highly accessible, reproducible, and interpretable tool for AI-powered ECG analysis, both for clinical application and future research.

4 Discussion

4.1 Summary and interpretation

ECG-XPLAIM is a deep learning model that aims to balance high diagnostic accuracy with interpretability in automated ECG analysis. Its Inception-style architecture, optimized for time-series data, employs multi-scale processing with adaptive receptive fields, allowing the detection of short-duration

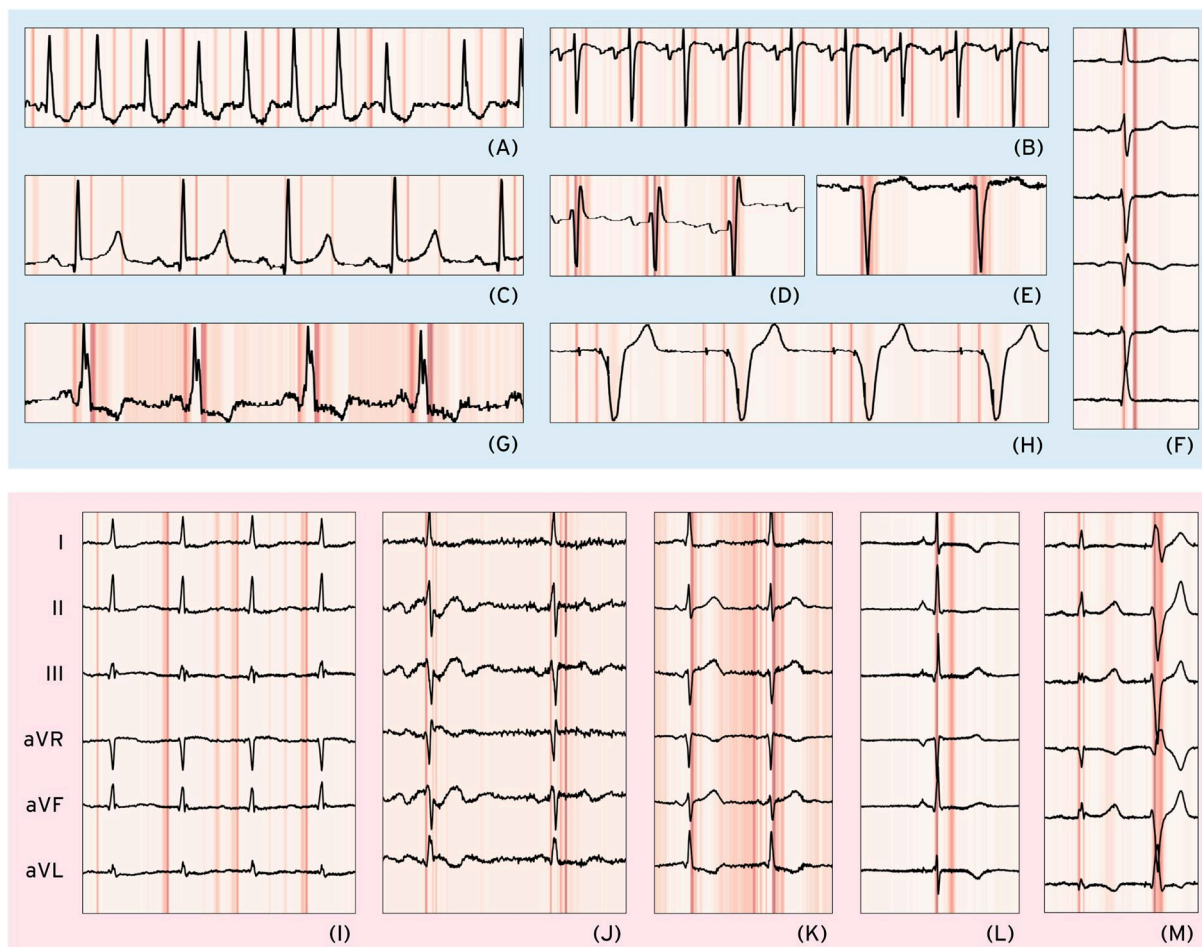


FIGURE 4

Selected correct and misclassified cases. Correctly classified cases (light blue background, upper half): (A) atrial fibrillation (AFib)—Lead II, (B) sinus tachycardia—Lead V1, (C) long QT (LQT)—Lead II, (D) right bundle branch block—Lead V1, (E) left bundle branch block—Lead V1, (F) left anterior fascicular block (LAFB)—Limb leads, (G) Wolff-Parkinson-White (WPW) pattern—Lead II, (H) paced rhythm (PACE)—Lead V1. Misclassified cases (light red background, lower half) with only limb leads shown: (I) AFib—false Positive due to coexisting first-degree AV block (I-AVB), possibly leading to P-wave misinterpretation, (J) LQT—false Negative possibly due to unclear/biphasic T-waves preventing accurate QT interval measurement, (K) WPW—False Negative where multiple points of interest before the QRS complex cause potential misinterpretation, (L) PACE—False Positive due to possible misclassification of a narrow QRS as a pacing spike, (M) LAFB—False Positive where a premature ventricular complex is possibly mistaken for a normal beat, leading to axis misinterpretation as LAFB-like.

waveform alterations, alongside global rhythm irregularities. Unlike traditional deep learning models that act as “black-box” classifiers, ECG-XPLAIM incorporates explainability mechanisms through one-dimensional Grad-CAM visualization, allowing for clinically meaningful interpretations and graspable explanations of its predictions.

Performance-wise, ECG-XPLAIM was assessed on held-out subsets of the development dataset (MIMIC-IV) for each task, where the model achieved metrics over 0.9 for all tasks. Most importantly, its diagnostic capability remained consistent on an external basis, when validated on the development-independent PTB-XL dataset. ECG-XPLAIM retained its competing performance, scoring metric values equal to or greater than 0.9 for most tasks, with only a few exceptions. Notably, certain arrhythmias, particularly LAFB, long QT, and WPW pattern, posed greater challenges in external evaluation, with sensitivity

dropping to approximately 0.7–0.8, while specificity and AUROC remained consistently high. Additional analyses confirmed that introducing conventional signal preprocessing (bandpass 0.5–40 Hz, notch 50 Hz) did not materially improve performance compared to training on raw signals. In fact, pacing detection sensitivity declined under filtering, likely because sharp pacemaker spikes were attenuated. These findings (Supplementary Table S3) support training on raw signals for the primary claims, while providing reproducible code for transparency.

This study acknowledges that class imbalance, particularly for rare arrhythmias like WPW and LQT, remains a persistent challenge in automated ECG interpretation. Imbalances may result in reduced sensitivity for these categories and affect generalization to broader patient populations. Systematic reviews of data augmentation and synthetic signal generation techniques

TABLE 3 Performance benchmarking of ECG-XPLAIM against baseline and external models.

Task	Sensitivity				Specificity				AUROC			
	ECG-XPLAIM	Vanilla CNN	GRU	External model**	ECG-XPLAIM	Vanilla CNN	GRU	External model	ECG-XPLAIM	Vanilla CNN	GRU	External model**
AFib	0.964 (0.944, 0.979)	0.644 (0.6, 0.686)*	0.966 (0.946, 0.98)	0.914 (0.886, 0.937)*	0.849 (0.835, 0.861)	0.837 (0.823, 0.85)*	0.83 (0.816, 0.844)*	0.919 (0.906, 0.931)*	0.958 (0.951, 0.965)	0.835 (0.823, 0.848)*	0.951 (0.943, 0.958)*	0.964 (0.956, 0.971)
STach	0.966 (0.946, 0.98)	0.892 (0.861, 0.918)*	0.954 (0.932, 0.971)	0.852 (0.818, 0.882)*	0.949 (0.941, 0.957)	0.935 (0.926, 0.944)*	0.894 (0.883, 0.905)*	0.98 (0.972, 0.985)*	0.983 (0.979, 0.988)	0.97 (0.965, 0.976)*	0.975 (0.97, 0.98)*	0.982 (0.976, 0.987)
RBBB	0.996 (0.986, 1)	0.986 (0.971, 0.994)	0.994 (0.983, 0.999)	0.922 (0.895, 0.944)*	0.936 (0.926, 0.944)	0.912 (0.901, 0.922)*	0.904 (0.892, 0.914)*	0.977 (0.969, 0.983)*	0.984 (0.98, 0.988)	0.975 (0.97, 0.98)*	0.98 (0.976, 0.985)	0.989 (0.985, 0.993)*
LBBB	0.988 (0.974, 0.996)	0.968 (0.949, 0.982)*	0.982 (0.966, 0.992)	0.93 (0.904, 0.951)*	0.894 (0.882, 0.905)	0.884 (0.872, 0.895)*	0.879 (0.866, 0.89)*	0.97 (0.962, 0.977)*	0.986 (0.983, 0.99)	0.979 (0.974, 0.984)*	0.985 (0.981, 0.989)*	0.985 (0.98, 0.99)
LAFB	0.694 (0.652, 0.734)	0.692 (0.649, 0.732)	0.692 (0.649, 0.732)	N/A	0.888 (0.876, 0.899)	0.862 (0.849, 0.874)*	0.883 (0.871, 0.894)	N/A	0.885 (0.874, 0.895)	0.882 (0.871, 0.892)	0.898 (0.888, 0.908)*	N/A
LQT	0.673 (0.577, 0.759)	0.7 (0.605, 0.784)	0.373 (0.282, 0.47)*	N/A	0.822 (0.809, 0.835)	0.612 (0.595, 0.629)*	0.841 (0.828, 0.853)*	N/A	0.81 (0.797, 0.823)	0.727 (0.713, 0.742)*	0.675 (0.66, 0.691)*	N/A
WPW	0.747 (0.633, 0.84)	0.347 (0.24, 0.465)*	0.187 (0.106, 0.293)*	N/A	0.901 (0.891, 0.911)	0.727 (0.711, 0.742)*	0.813 (0.799, 0.826)*	N/A	0.863 (0.852, 0.874)	0.562 (0.546, 0.579)*	0.521 (0.504, 0.537)*	N/A
PACE	0.959 (0.929, 0.978)	0.962 (0.933, 0.981)	0.969 (0.942, 0.986)	N/A	0.977 (0.972, 0.982)	0.825 (0.811, 0.838)*	0.682 (0.665, 0.698)*	N/A	0.985 (0.981, 0.989)	0.977 (0.972, 0.982)*	0.977 (0.972, 0.982)*	N/A

AFib, atrial fibrillation; AUROC, area under the receiver operating characteristic; CNN, convolutional neural network; GRU, gated recurrent unit; LAFB, left anterior fascicular block; LBBB, left bundle branch block; LQT, long QT; N/A, not available; PACE, paced rhythm; RBBB, right bundle branch block; STach, sinus tachycardia; WPW, Wolff-Parkinson-White pattern. Metrics are reported with 95% confidence intervals; Values marked with an asterisk (*) indicate significant differences at 0.05 level, as compared to ECG-XPLAIM; ECG-XPLAIM was benchmarked against CNN and GRU across all eight classification labels, while the external model (27) was evaluated only on the four common labels supported by both models (AFib, STach, RBBB, LBBB). For each comparison, at least 500 records were sampled per label, except LQT (110), WPW (75) and PACE (290) due to limited data availability (more details in [Supplementary Material](#)).

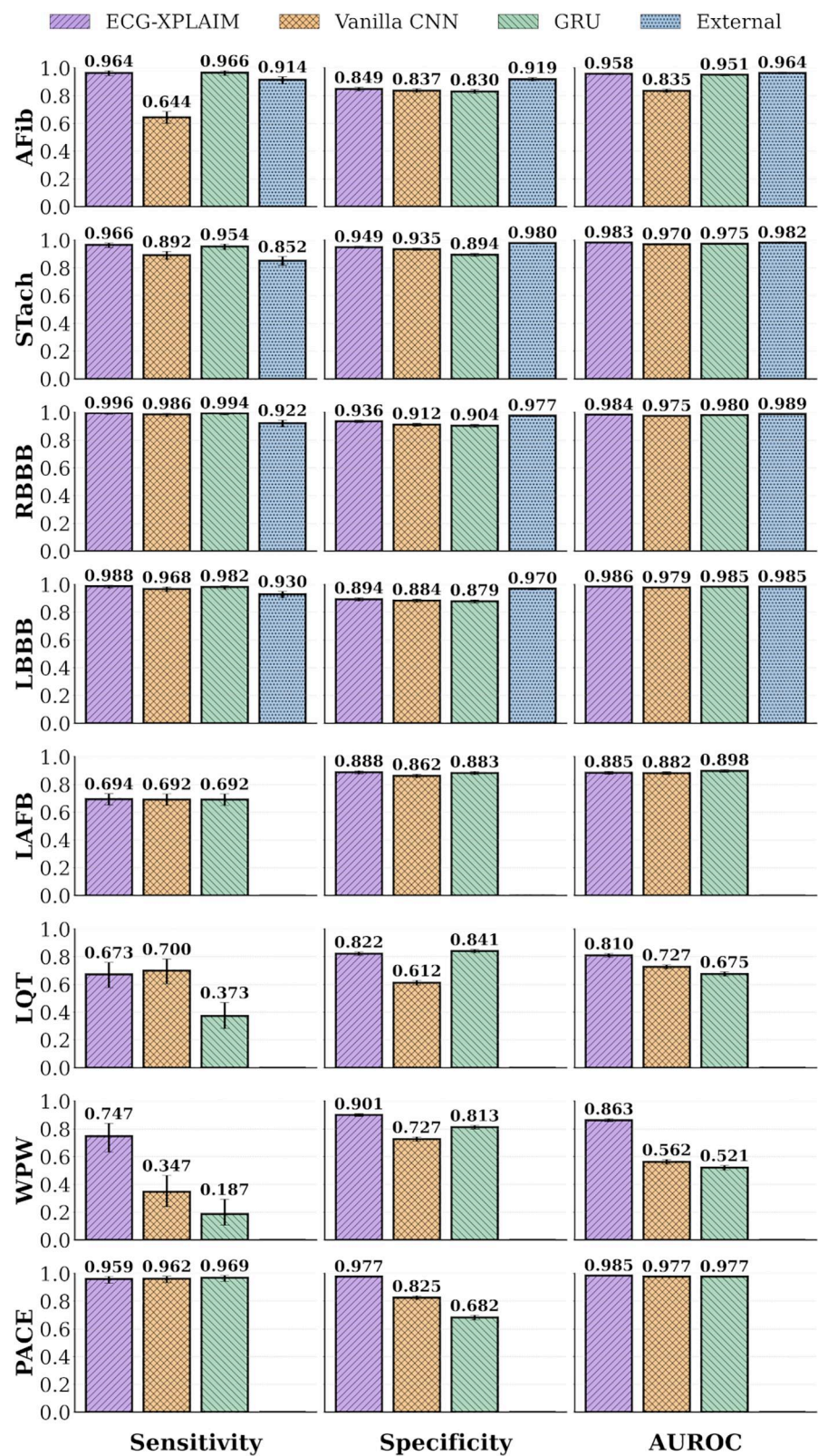


FIGURE 5
Performance bar graphs for specific arrhythmias. Classification performance metrics, including sensitivity, specificity, and area under the receiver operating characteristic (AUROC) curve, are displayed for individual arrhythmias. AFib, atrial fibrillation; CNN, convolutional neural network; External, external, pre-trained model (27); GRU, gated recurrent unit; LAFB, left anterior fascicular block; LBBB, left bundle branch block; LQT, long QT; PACE, paced rhythm; RBBB, right bundle branch block; STach, sinus tachycardia; WPW, Wolff-Parkinson-White pattern.

suggest that targeted strategies can ameliorate the impact of imbalance while enhancing the robustness and fairness of diagnostic models (31). To mitigate this, the model training employed a maximum cap of 50,000 samples per class where feasible, ensuring balanced representation across major arrhythmia categories. For WPW detection, due to data scarcity, a positive-to-negative sample ratio of approximately 1:12 (600 positive vs. 1,220 negative samples) was maintained, while LQT classification utilized around 39,000 samples per class. Other categories such as atrial fibrillation, sinus tachycardia, and conduction disturbances had similarly controlled sample sizes, with negative classes including non-target arrhythmic pathologies to reflect real-world complexity. These efforts aimed to limit dominant class bias without aggressive oversampling or augmentation. Supplementary analyses suggest that targeted augmentation and transfer learning strategies could further enhance detection of rare arrhythmias, supporting ongoing research in this direction (31).

The performance discrepancy between external and internal validation on specific labels, may stem from training set biases, class underrepresentation, or even inherent model limitations in capturing the subtle waveform characteristics. In particular, the reduced sensitivity for LQT (0.691) and WPW (0.773) in PTB-XL can be explained by cross-dataset domain shifts (differences in lead placement, sampling rates, acquisition chains, and annotation criteria), phenotype definition heterogeneity (e.g., QT correction formulas), and waveform ambiguity (e.g., when borderline QT prolongation or delta-like pre-excitation morphologies overlap with ectopy). These considerations highlight the importance of dataset-specific calibration and may motivate future domain adaptation strategies. Data augmentation (31–33), targeted fine-tuning (34), and signal pre-processing (35), might help enhance the detection of these patterns. Illustrative fine-tuning experiments on PTB-XL (0%–50% train/test splits) demonstrated that features transfer sufficiently across datasets: WPW AUROC improved steadily with additional PTB-XL data, LQT sensitivity spiked with very small fractions but at the expense of specificity, and TACHY/CD/PACE performance remained relatively robust. AUROC generally increased as more external data were used (Supplementary Table S5, Supplementary Figure S3). We emphasize that this violates strict external independence, so it is not part of our primary claims, but it highlights transfer learning as a promising future avenue.

Benchmarking against three counterpart models—a vanilla CNN, a more advanced two-layer GRU model, and an external, pre-trained ResNet-based architecture—revealed that ECG-XPLAIM generally demonstrates competitive performance. ECG-XPLAIM consistently outperformed both baseline CNN and GRU models, achieving higher sensitivity, specificity, and AUROC scores across most tasks. Compared to the external, pre-trained model, ECG-XPLAIM demonstrated higher sensitivity, leading to a slightly reduced specificity. This preference for minimizing false negatives aligns with its potential role as a screening tool, where missing critical arrhythmias is more concerning than erroneously flagging some

normal cases. Benchmarking against such leading architectures demonstrates the competitive diagnostic performance of the present approach. Importantly, comparison with large-scale neural network frameworks for ECG interpretation, such as that developed by Ribeiro et al., highlights the added clinical value of integrated interpretability, which is increasingly recognized as a prerequisite for real-world deployment in cardiology settings (27).

4.2 Clinical applicability and added value

ECG-XPLAIM is designed to integrate seamlessly into clinical workflows by prioritizing both diagnostic accuracy and interpretability. Its high sensitivity ensures that clinically significant arrhythmias are detected early, aiding in timely referrals to cardiology specialists and reducing the risk of underdiagnosis in severe conditions such as conduction blocks or arrhythmias predisposing to ventricular events. Additionally, it enhances efficiency in high-throughput diagnostic environments by assisting in automated triaging of abnormal ECGs, reducing the burden on specialists, offering a fatigue-free screening solution, and facilitating early identification of high-risk patients.

Unlike many prior approaches that primarily distinguish normal from abnormal ECGs, ECG-XPLAIM focuses on challenging arrhythmias and leverages multi-class, overlap-tolerant training. Negative classes in each task are not purely “normal” recordings but instead exclude only the target arrhythmic entity while potentially containing other conditions. Consequently, ECG-XPLAIM learns to differentiate subtle, overlapping abnormalities—a skill crucial in real-world practice where arrhythmias often coexist or mimic one another. Furthermore, ECG-XPLAIM was specifically trained on well-defined classification tasks, ensuring a balanced representation across involved classes and focusing only on arrhythmias that are challenging to differentiate. This task-specific design increases its applicability in real-world settings, where ECGs often present overlapping abnormalities that require fine-grained discrimination.

ECG-XPLAIM represents a highly scalable solution, capable of handling large-scale ECG datasets, supporting both high-volume batch processing and low-latency real-time inference, making it suitable for both retrospective research and live clinical deployments. From a technical perspective, it exhibits short inference times (4.5–16 milliseconds per 10-s ECG), enabling near real-time deployment in edge- or server-based infrastructures. The model is open-source, enabling research groups to extend its architecture, adapt it for novel classification tasks, and implement custom modifications tailored to specific clinical needs. The availability of pre-trained weights facilitates direct deployment without extensive retraining, while also allowing for transfer learning and fine-tuning on new datasets, significantly reducing computational costs and making it accessible to a broader user base.

An additional threshold analysis (Supplementary Table S6) demonstrated how sensitivity and specificity trade-offs evolve across thresholds 0.0–1.0. For example, TACHY achieved a balanced operating point at 0.4–0.5 (SEN 0.968–0.955, SPE

0.952–0.969), CD balanced at 0.5, and PACE maintained excellent performance across a wide range of thresholds. In contrast, LQT sensitivity dropped steeply as thresholds increased, suggesting that lower thresholds (≤ 0.3 –0.4) may be preferable in screening contexts, but not for disease confirmation, due to high risk for false positives. These findings support the potential for site-specific threshold personalization, where operating points can be pre-specified depending on whether the model is used for broad screening (favoring sensitivity) or confirmatory diagnostics (favoring specificity).

More broadly, ECG-XPLAIM fits into the growing role of AI in electrophysiology (EP) workflows, which extends beyond ECG classification into procedural guidance, ablation planning, and arrhythmia risk stratification (36). AI tools are increasingly applied for automated mapping of atrial and ventricular arrhythmias, predicting catheter ablation outcomes, optimizing device programming, and guiding individualized risk assessment for sudden cardiac death. Within this framework, ECG-based algorithms such as ours provide the critical front-end signal interpretation layer: by ensuring reliable, explainable detection of arrhythmias and conduction disturbances, ECG-XPLAIM can serve as the entry point that feeds into downstream EP workflows, including rhythm monitoring, decision-support in invasive procedures, and integration with longitudinal risk prediction models. In this way, ECG-XPLAIM's emphasis on transparency and adaptability positions it as a foundational component in the translational pipeline of AI in EP. Explainable and open-source models such as ours align with these translational goals, as highlighted in recent consensus perspectives on AI in EP (13).

4.3 Explainability and trust in AI-driven diagnostics

A primary barrier to the clinical adoption of deep learning in healthcare is the “black-box” nature of most models. The interpretability of deep learning models applied to electrocardiogram analysis is increasingly acknowledged as essential for facilitating clinical adoption and patient safety. Contemporary literature highlights that XAI methods require rigorous evaluation to ensure reliability and relevance in clinical practice. Salih et al. conducted a systematic review of XAI evaluations in cardiology, revealing that only a minority of studies applied systematic assessment: 37% benchmarked XAI quality based on prior literature, 11% involved clinicians as domain experts, and 11% relied on quantitative proxies or statistical analysis, while 43% did not assess explanation quality at all. The authors advocate for formal, multi-dimensional frameworks that include faithfulness, fidelity, and direct clinician feedback, emphasizing that thorough evaluation of explanations is critical for the development of trustworthy and safe AI models in medicine (14).

Recent advances also underscore the utility of combining multiple interpretability techniques and actively involving clinicians in validation processes. Zhang et al. demonstrated the

application of Grad-CAM in medical text classification, illustrating how visualization of salient features through heatmaps can intuitively communicate the basis for predictions to human users. Their comparative study using word embeddings and various classifier architectures (Word2Vec, BERT, ResNet, CNN, Bi-LSTM) showed that integrating Grad-CAM with high-performing deep learning models enables more transparent identification of decision-influencing input signals. The study found that Grad-CAM visualization reliably highlighted text regions most relevant to the model's outputs, supporting the practical integration of XAI in clinical decision making and error analysis (37).

ECG-XPLAIM confronts this challenge by integrating Grad-CAM-based explainability, enabling visualization of the waveform regions that contribute most to its predictions. This transparency fulfills several objectives. First, it strengthens clinician trust by revealing the model's decision-making process. Second, it facilitates potential feature discovery, unearthing subtle waveform variations that may carry clinical significance. Third, it supports adherence to emerging regulatory guidelines—such as those from the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA)—that increasingly emphasize interpretability requirements for medical AI systems (38, 39). Finally, it assists in error analysis: highlighting waveforms that led to misclassifications allows targeted improvements to the model's training and architecture.

In this study, Grad-CAM visualizations helped pinpoint areas of interest, in both correctly identified and misclassified cases. For AFib, the model consistently focused on the absence of P waves in the pre-QRS region (Figure 4A), while the STach detection was primarily driven by P-wave presence and regularly appearing points of interest that signify rhythmicity (Figure 4B). RBBB and LBBB cases showed strong attention to the QRS complex morphology (Figures 4D,E), while in LAFB, the model seemed to capture the associated axis deviation-related changes (Figure 4F). For LQT, ECG-XPLAIM correctly identified the QT interval by focusing on the onset and termination of the repolarization phase in certain beats (Figure 4C). WPW classification relied on the characteristic delta wave and the PR interval (Figure 4G). Paced rhythm cases were accurately identified by highlighting both atrial and ventricular pacing spikes across all beats (Figure 4H).

On the other hand, false classifications revealed cases where the model's attention was misdirected, particularly for arrhythmias with lower sensitivities. In an AFib false positive case, ECG-XPLAIM incorrectly interpreted a conduction delay due to first-degree AV block (I-AVB) as an absent P wave (Figure 4I), demonstrating a potential bias in P-wave localization. Conversely, in a false negative STach case, the model correctly detected rapid rhythm but misclassified it as AFib due to near-fusion of the P wave with the preceding T wave at high heart rates. Bundle branch block misclassifications were primarily linked to variations in QRS duration that seemed borderline. False positive classifications often involved misinterpretation of extrasystoles or morphological changes in the QRS complex (Figure 4M). For

LQT false negatives, the model's attention was restricted to a segment within the repolarization phase rather than spanning the full QT interval, for some cases (Figure 4J). In WPW false positive examples, ECG-XPLAIM placed significant focus on the pre-QRS regions of wide-QRS extrasystoles, mistaking premature beats for delta waves, indicating a bias in distinguishing abnormal conduction patterns. Similarly, a false positive classification of pacing occurred when the model misinterpreted a narrow QRS complex as a pacing spike (Figure 4L). (Detailed examples with explanations are offered in Supplementary Section S8.) These findings indicate the origin of prediction faults and suggest strategies to mitigate them, such as augmenting training data with borderline and atypical presentations (31).

Despite its utility, Grad-CAM is not a perfect solution. Although it highlights influential waveform regions, it does not fully elucidate the underlying rationale—why certain features are attributed to one arrhythmia rather than another (37). Future research could explore more advanced or complementary explainable AI techniques, potentially integrating rule-based logic or interpretability frameworks that capture inter-lead relationships. These refinements may further reduce model rigidity and enhance its ability to handle the complexities of real-world ECG data.

4.4 Limitations and future directions

Despite strong performance and explainability, ECG-XPLAIM faces certain limitations. Class imbalance was present across tasks, particularly for rare arrhythmias such as WPW and LQT, where positive samples were substantially fewer than negatives (e.g., WPW 600 vs. 1,220). Although we capped maximum samples per class and applied balanced mini-batching, residual imbalance may have contributed to lower external sensitivities. Future work could leverage more targeted data collection or dedicated augmentation strategies for ECG signals to enrich rare classes and improve model calibration (31). Importantly, the reliability of diagnostic labels was confirmed via inter-observer agreement analysis: in a random 10% subset of MIMIC-IV, kappa values ranged from 0.990 to 0.998 across all labels, supporting the sufficiency of pre-annotations (Supplementary Table S4). Finally, real-world clinical performance can only be validated prospectively; although robust, retrospective testing on MIMIC-IV and PTB-XL does not guarantee identical outcomes in diverse clinical environments.

While Grad-CAM visualizations partly address the interpretability gap, they do not provide an explicit rationale for how certain features lead to a diagnosis. For instance, identifying a lengthened QRS complex does not clarify how the model distinguishes between RBBB and LBBB. More sophisticated XAI methods could further demystify the decision process and illuminate nuanced inter-lead relationships that underlie arrhythmia detection.

Finally, exploratory transfer learning experiments on PTB-XL (Section 6 of the Supplementary Material) demonstrated that ECG-XPLAIM's feature representations are transferable across

datasets, particularly for WPW and LQT detection. While these experiments can improve metrics, such as AUROC and sensitivity, for specific labels under certain splits, they are illustrative only, as they break the requirement for independence in external validation. Nonetheless, they motivate future research directions.

Future research will focus on refining ECG-XPLAIM's generalization and interpretability. Fine-tuning on localized, hospital-specific datasets could account for regional ECG variations and acquisition protocols, while federated learning approaches may broaden the model's adaptability without centralized data pooling (40). Investigating additional XAI techniques or combining Grad-CAM with rule-based logic could strengthen interpretability and expedite regulatory acceptance. Finally, prospective clinical trials will be essential to evaluate real-world feasibility, confirm performance in diverse patient populations, and measure clinical outcomes and workflow improvements attributable to ECG-XPLAIM's integration.

5 Conclusions

In this work, we propose ECG-XPLAIM, an explainable deep learning model for automated arrhythmia detection, which demonstrates robust generalization in both the MIMIC-IV and PTB-XL datasets. ECG-XPLAIM outperforms baseline CNN and advanced GRU models in most classification tasks and offers performance comparable to a state-of-the-art pre-trained network, with a priority to minimize the risk of underdiagnosis. By emphasizing sensitivity, it reduces missed diagnoses, making it particularly well-suited for screening workflows. Its integrated Grad-CAM mechanism provides interpretable visualizations of the waveform regions guiding classification, simultaneously furnishing critical feedback for model refinement. Additional analyses confirmed the robustness of training on raw signals, and the flexibility to adapt performance through threshold calibration or transfer learning to new datasets. These features strengthen the model's translational potential. Although further optimization for rare arrhythmias, larger datasets, and real-world prospective validation are warranted, ECG-XPLAIM's scalability, open-source implementation, and rapid inference, position it as a valuable tool for integrating AI-driven cardiac diagnostics into clinical practice.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://physionet.org/>. The ECG-XPLAIM source code, along with use instructions, are available on GitHub at <https://github.com/ppantele/ECG-XPLAIM>. The trained weights of ECG-XPLAIM for each task are available on Zenodo at <https://zenodo.org/records/14968732>.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

PP: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft. SR-C: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft. JV: Formal analysis, Methodology, Resources, Supervision, Writing – review & editing. AA: Conceptualization, Formal analysis, Resources, Software, Supervision, Writing – review & editing. IG: Formal analysis, Writing – original draft. GZ: Formal analysis, Methodology, Writing – original draft. MS: Formal analysis, Validation, Writing – original draft. PD: Data curation, Investigation, Resources, Validation, Writing – original draft. JM: Data curation, Methodology, Resources, Writing – review & editing. TP: Data curation, Investigation, Methodology, Resources, Validation, Writing – review & editing. EO: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing. GS: Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research protocol was partially funded by the Hellenic Society of Cardiology.

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. *Circulation*. (2022) 145:e153–639. doi: 10.1161/CIR.0000000000001052
2. Zeppenfeld K, Tfelt-Hansen J, De Riva M, Winkel BG, Behr ER, Blom NA, et al. 2022 ESC guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. *Eur Heart J*. (2022) 43:3997–4126. doi: 10.1093/eurheartj/ehac262
3. Salerno SM, Alguire PC, Waxman HS. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. *Ann Intern Med*. (2003) 138:751–60. doi: 10.7326/0003-4819-138-9-200305060-00013
4. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. (2019) 25:65–9. doi: 10.1038/s41591-018-0268-3
5. Kolk MZH, Ruipérez-Campillo S, Alvarez-Florez L, Deb B, Bekkers EJ, Allaart CP, et al. Dynamic prediction of malignant ventricular arrhythmias using neural

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2025.1659971/full#supplementary-material>

networks in patients with an implantable cardioverter-defibrillator. *eBioMedicine*. (2024) 99:104937. doi: 10.1016/j.ebiom.2023.104937

6. Pantelidis P, Oikonomou E, Gialamas I, Goliopoulou A, Sarantos S, Zakyntinos GE, et al. Decoding the heart: how artificial intelligence is transforming cardiology. *J Med Artif Intell*. (2025) 8:9. doi: 10.21037/jmai-24-139

7. Dhingra LS, Aminorroaya A, Sangha V, Pedrosa AF, Asselbergs FW, Brant LCC, et al. Heart failure risk stratification using artificial intelligence applied to electrocardiogram images: a multinational study. *Eur Heart J*. (2025) 46:1044–53. doi: 10.1093/eurheartj/ehae914

8. Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017–2023. *Front Physiol*. (2023) 14:1246746. doi: 10.3389/fphys.2023.1246746

9. Pantelidis P, Oikonomou E, Souvaliotis N, Spartalis M, Bampa M, Papapetrou P, et al. Optimising and validating deep learning approaches for diagnosing atrial fibrillation from few-lead ambulatory electrocardiogram signals. *Europace*. (2022) 24:euac053-561. doi: 10.1093/europace/euac053.561

10. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. (2018) 15:20170387. doi: 10.1098/rsif.2017.0387
11. Oikonomou EK, Khera R. Designing medical artificial intelligence systems for global use: focus on interoperability, scalability, and accessibility. *Hellenic J Cardiol*. (2025) 81:9–17. doi: 10.1016/j.hjc.2024.07.003
12. Kolk MZH, Ruipérez-Campillo S, Allaert CP, Wilde AAM, Knops RE, Narayan SM, et al. Multimodal explainable artificial intelligence identifies patients with non-ischaemic cardiomyopathy at risk of lethal ventricular arrhythmias. *Sci Rep*. (2024) 14:14889. doi: 10.1038/s41598-024-65357-x
13. Svennberg E, Han JK, Caiani EG, Engelhardt S, Ernst S, Friedman P, et al. State of the art of artificial intelligence in clinical electrophysiology in 2025: a scientific statement of the European Heart Rhythm Association (EHRA) of the ESC, the Heart Rhythm Society (HRS), and the ESC Working Group on E-cardiology. *Europace*. (2025) 27:eua071. doi: 10.1093/europace/euaf071
14. Salih AM, Galazzo IB, Gkontra P, Rauseo E, Lee AM, Lekadir K, et al. A review of evaluation approaches for explainable AI with applications in cardiology. *Artif Intell Rev*. (2024) 57:240. doi: 10.1007/s10462-024-10852-w
15. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *2017 IEEE International Conference on Computer Vision (ICCV)*; Venice: IEEE (2017). p. 618–26. doi: 10.1109/ICCV.2017.74
16. Pantelidis P, Oikonomou E, Lampsas S, Souvaliotis N, Spartalis M, Vavuranakis M, et al. Inside the “brain” of an artificial neural network: an interpretable deep learning approach to paroxysmal atrial fibrillation diagnosis from electrocardiogram signals during sinus rhythm. *Eur Heart J Digit Health*. (2022) 3:ztac076-2781. doi: 10.1093/eurheartj/ehac544.2781
17. Ismail Fawaz H, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, et al. Inceptiontime: finding AlexNet for time series classification. *Data Min Knowl Discov*. (2020) 34:1936–62. doi: 10.1007/s10618-020-00710-y
18. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. (2023) 10(1):1. doi: 10.1038/s41597-022-01899-x
19. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov P, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for Complex physiologic signals. *Circulation*. (2000) 101:e215–20. doi: 10.1161/01.CIR.101.23.e215
20. Wagner P, Ströthoff N, Bousselot R-D, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data*. (2020) 7:154. doi: 10.1038/s41597-020-0495-6
21. Yoo J, Yoo I, Youn I, Kim S-M, Yu R, Kim K, et al. Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability. *Comput Methods Programs Biomed*. (2022) 226:107079. doi: 10.1016/j.cmpb.2022.107079
22. Collins GS, De Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. (2014) 14:40. doi: 10.1186/1471-2288-14-40
23. Kingma DP, Ba J. Adam: a method for stochastic optimization (2017). doi: 10.48550/arXiv.1412.6980
24. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. (1934) 26:404–13. doi: 10.1093/biomet/26.4.404
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837. doi: 10.2307/2531595
26. Ebrahimi Z, Loni M, Daneshdab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst Appl X*. (2020) 7:100033. doi: 10.1016/j.eswax.2020.100033
27. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. (2020) 11:1760. doi: 10.1038/s41467-020-15432-4
28. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. (1947) 12:153–7. doi: 10.1007/BF02295996
29. Hawass NE. Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. *Br J Radiol*. (1997) 70:360–6. doi: 10.1259/bjr.70.832.9166071
30. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall (1993).
31. Rahman MM, Rivolta MW, Badilini F, Sassi R. A systematic survey of data augmentation of ECG signals for AI applications. *Sensors*. (2023) 23:5237. doi: 10.3390/s23115237
32. Pantelidis P, Oikonomou E, Bampa M, Gialamas I, Goliopoulou A, Mystakidi VC, et al. Generative adversarial networks (GANs) to produce synthetic 12-lead electrocardiogram signals for specific and rare diseases: a novel, powerful tool towards clinical advancements. *Eur Heart J*. (2024) 45:ehae666.3427. doi: 10.1093/eurheartj/ehae666.3427
33. Ruipérez-Campillo S, Ryser A, Sutter TM, Feng R, Ganesan P, Deb B, et al. Can generative AI learn physiological waveform morphologies? A study on denoising intracardiac signals in ischemic cardiomyopathy. *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; Orlando, FL, USA: IEEE (2024). p. 1–4. doi: 10.1109/EMBC53108.2024.10782966
34. Weimann K, Conrad TOF. Transfer learning for ECG classification. *Sci Rep*. (2021) 11:5251. doi: 10.1038/s41598-021-84374-8
35. Safdar MF, Nowak RM, Palka P. Pre-processing techniques and artificial intelligence algorithms for electrocardiogram (ECG) signals analysis: a comprehensive review. *Comput Biol Med*. (2024) 170:107908. doi: 10.1016/j.combiomed.2023.107908
36. Cersosimo A, Zito E, Pierucci N, Matteucci A, La Fazio VM. A talk with ChatGPT: the role of artificial intelligence in shaping the future of cardiology and electrophysiology. *J Pers Med*. (2025) 15:205. doi: 10.3390/jpm15050205
37. Zhang H, Ogasawara K. Grad-CAM-Based explainable artificial intelligence related to medical text processing. *Bioengineering*. (2023) 10:1070. doi: 10.3390/bioengineering10091070
38. U.S. Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices (2024). Available online at: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-ml-enabled-medical-devices> (Accessed June 25, 2025).
39. European Medicines Agency. Artificial intelligence—EMA (2024). Available online at: <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/artificial-intelligence> (Accessed June 25, 2025).
40. Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. *Cell Rep Med*. (2024) 5:101419. doi: 10.1016/j.xcrm.2024.101419

Frontiers in Cardiovascular Medicine

Innovations and improvements in cardiovascular treatment and practice

Focuses on research that challenges the status quo of cardiovascular care, or facilitates the translation of advances into new therapies and diagnostic tools.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Cardiovascular Medicine

