Demystifying AI for early AF detection: enhancing diagnostic transparency across modalities

Justin Lee Mifsud, Mark Adrian Sammut, Claire Galea

Authors

Justin Lee Mifsud
Academic

Claire Galea
Student Nurse

Faculty of Health Sciences, University of Malta, Msida, Malta

Mark Adrian Sammut
Consultant Electrophysiologist
Mater Dei Hospital and Faculty of Medicine and Surgery, University of Malta, Msida, Malta

Correspondence to:
Dr J L Mifsud
(justin-lee.mifsud@um.edu.mt)

Key words

artificial intelligence, atrial fibrillation, electrocardiography

First published online 5th November 2025

doi: 10.5837/bjc.2025.051

Br J Cardiol 2025;**32**(4)

This article explores using artificial intelligence (AI) to detect atrial fibrillation (AF) early, highlighting its potential to revolutionise cardiology. It reviews numerous studies demonstrating AI's superior accuracy to traditional methods, particularly in leveraging electrocardiography data from various sources like smart devices and chest radiographs. A key concern addressed is the 'black box' nature of some AI algorithms, emphasising the critical need for transparency to build clinician confidence and ensure ethical patient care. It concludes by advocating for policy changes and further research to enhance AI algorithm transparency and integration into clinical practice.

Introduction

Despite advances in managing atrial fibrillation (AF), it remains a major contributor to cardiovascular morbidity and mortality,[1] placing a significant burden on both public health costs and the healthcare system.[1] Cardiology is at the forefront of the artificial intelligence (AI) revolution within medicine, integrating AI with traditional diagnostic methods for timely interventions. For example, an AI-driven tool is the 10-second AI-enabled electrocardiogram (ECG), which could detect or even predict AF in patients who may have otherwise gone undiagnosed at the point of care.[2–4] By identifying AF earlier, this technology has the potential to reduce AF-related complications, such as ischaemic stroke, heart failure, sudden death, and other cardiovascular events.[5] Over recent years, researchers have evaluated various AI techniques, including convolutional neural networks (CNNs) and deep neural networks (DNNs), both demonstrating promising outcomes. However, a significant limitation of AI is the 'black box' problem, where the lack of transparency in AI algorithms' inner workings poses practical and ethical challenges. **Table 1** answers common questions regarding its applications, effectiveness and limitations.

Traditional versus AI-based AF diagnostic approaches

While ECG can be a diagnostic tool in healthcare settings, the standard 60-second ECG strips may exhibit a normal sinus rhythm (NSR) during evaluation, potentially overlooking an episode of AF and leading to delayed diagnosis. For example, this scenario was observed in a study assessing the efficacy of opportunistic screening in public healthcare settings, where no significant increase in AF detection was noted compared with standard care. Consequently, traditional opportunistic screening for AF in public healthcare settings may not be deemed clinically viable.[6]

Unlike the traditional screening approach, AI-guided AF screening for patients at increased risk of stroke, but no known AF, offers a practical, patient-centred, and scalable solution to reduce unnecessary strain on the healthcare system, as it has a higher detection rate of AF than traditional methods (detection rate, AI-guided 10.6% *vs.* usual care 3.6%, $p<0.0001$).[3]

Recent studies highlight the growing role of AI in AF detection and diagnosis across various modalities. For example, combining a deep-learning (DL) model with chest radiography showed that indicators of AF are visible, even on static images, offering radiologists an additional method for detecting AF. This approach operates efficiently with a low computational burden and utilises smaller file sizes, unlike standard digital chest radiographs, enabling deployment on commonly available hardware.[7]

In a retrospective cohort study, a CNN technique was used to encode seven-second ECG segments into a latent space by training it to detect paroxysmal AF at a segment level. Then, 24-hour ECG sequences

Table 1. Detecting atrial fibrillation (AF) with artificial intelligence (AI): key questions and answers

Question	Answer
How does AI contribute to AF detection?	AI algorithms, particularly those integrated with electrocardiography (ECG) and wearable technology, offer a more efficient and accurate analysis than traditional methods, allowing for earlier identification of AF even in individuals displaying normal sinus rhythm (NSR) on standard ECGs
What are the advantages of AI-powered AF detection over conventional methods?	AI excels in analysing vast datasets, identifying subtle patterns indicative of AF that might be missed by human observation, improving diagnostic accuracy and leading to earlier detection and better patient outcomes. AI also facilitates large-scale screenings and risk stratification
What are the potential benefits of AI-based AF detection for patients?	Early AF detection through AI empowers patients with the knowledge to make informed decisions about their health. It facilitates proactive management of AF, potentially reducing the risk of serious complications and improving long-term quality of life
How does the increasing prevalence of smart devices contribute to AI-driven AF detection?	The widespread adoption of smart devices equipped with sensors and AI algorithms presents a unique opportunity for continuous monitoring and early AF detection. These devices can analyse real-time data, detect irregular heart rhythms, and alert users and healthcare providers for timely intervention
Are there any limitations to using AI for AF detection?	A significant concern about AI is the 'black box' problem, where the algorithm's decision-making process remains opaque. This lack of transparency raises ethical concerns and hinders clinician confidence in AI-derived diagnoses
How can the 'black box' problem be addressed to enhance trust in AI?	Techniques like gradient-weighted class activation mapping (Grad-CAM) can be implemented to visualise the ECG sections crucial for AI's AF prediction. Therefore, it can foster transparency, allow clinicians to understand the rationale behind AI decisions, and facilitate discussions with patients about their diagnoses
What implications does AI hold for the future of cardiology practice?	AI is poised to transform cardiology by enabling more precise and timely AF detection, leading to personalised interventions and improved patient care

were mapped into this space, and a gradient-boosting machine (GBM) was trained to analyse patterns across the full Holter recording. This approach captures segment-level features and time-related patterns within a patient's ECG. This CNN-based AI model for Holter monitoring performed optimally, avoiding the supraventricular ectopy burden, analysing complete 24-hour recordings, and focusing on night-time data.[8] Other researchers have used the wavelets technique for signal processing and feature extraction. The wavelet technique is a mathematical tool that extracts meaningful features, reduces noise, and analyses in multi-scale before the CNN processes the data and feeds it into an AI model. When integrated with a CNN technique, wavelet achieved the highest accuracy,[9,10] predicting AF from a single-lead ECG exhibiting NSR. This wavelet–CNN combination demonstrated enhanced performance across a broader age range, showcasing its effectiveness in AF prediction.[4,11]

In wearable and mobile devices equipped with AI for detecting arrhythmia, a systematic review and meta-analysis demonstrated the superiority of DNN-based over traditional machine-learning algorithms.[12] Unlike traditional machine learning, which requires manual feature extraction, DNNs use multiple layers to process data, learn patterns, and automatically make predictions. Other researchers have used advanced CNN, combining photoplethysmography, ECG and an AF-identifying AI algorithm. This CNN-based AI model has shown potential for detecting AF. The device equipped with this AI model is the Amazfit Health Band 1S, which has been shown to effectively gather ECG data, even in environments with signal interference, and detect rhythms, such as premature beats.[13]

In risk stratification for detecting non-persistent AF, gradient-weighted class activation mapping (Grad-CAM) was developed to explain the decision-making of an AI model to detect AF promptly within short recording windows (one week). The Grad-CAM highlights the ECG sections responsible for predicting AF, further enhancing prediction interpretability for clinical applications.[14] Similarly, a DL model capable of high-performance AF diagnosis with decision-making transparency has been developed using a neural-backed ensemble tree (NBET).[15]

Although AI seems to be a feasible fit for many clinical facilities, it may not be applicable in every clinical setting; for example, an AI model trained using the AI-ECG AF algorithm showed limited clinical utility in patients presenting with palpitations at the emergency department. However, this lack of utility could be due to differences in the mean age between the sample and the original population used in model creation.[16] This highlights the critical importance of training data in AI development, ensuring that the data, features and weights are comprehensive and accurately designed to avoid deficiencies in the model's training.

Challenges with using AI in cardiac clinical practice

As AI advances in supporting diagnostic accuracy, there is a need to emphasise the importance of transparency in AI algorithms, enabling clinicians to be confident with AI-derived diagnoses. The lack of transparency in AI algorithms, often called the 'black box' problem, represents a significant research gap in AI. In the clinical setting, the best medical practice involves shared decision-making between clinicians and patients, guided by ethical principles, such as beneficence, autonomy, and respect. Without a clear rationale for decisions regarding patient health, the foundation for treatment remains unsubstantiated. Unless further research is conducted on the justifiability and transparency of AI algorithms, the ethical barrier posed by the AI 'black box' will persist.[17]

There could also be a question of equity. If

AI is mainly trained on data imbalance, such as having more affluent populations, it might completely overlook subtle signs or risk factors that are more common in underserved communities, leading to inaccurate risk assessments and, potentially, delayed diagnoses. It is essential to use diverse and representative data sets when training these algorithms, and to ensure that all population segments are reflected to avoid unintentional bias, including ensuring that diverse divisions are involved in developing and implementing these technologies. It is necessary to gather perspectives from different communities, including clinicians, patients, and ethicists, to safeguard these AI systems and ensure they are sensitive to everyone's needs and concerns. Beyond data diversity, the issue of access should be carefully addressed. If these powerful AI tools are only available to people who can afford them, there is a risk of creating a two-tiered healthcare system where the wealthy benefit disproportionately. Therefore, it becomes a social justice issue, as much as a technological one. It is a challenge that needs a joint effort from policymakers, healthcare providers, and even tech developers, to ensure that everyone can benefit from these advances in AI-driven healthcare, not just a select few.

Enhancing transparency in AI algorithms: clinician and patient benefits

The limitations inherent to the 'black box' problem, reflecting limited algorithm transparency and justifiability, could be mitigated by integrating Grad-CAM with any chosen AI modality, enhancing its interpretability and transparency. Grad-CAM could benefit clinicians by providing a heat map of an ECG, highlighting precisely which parts the AI algorithm focused on to reach its diagnosis. A detailed figure illustrating this concept is available in Kim *et al.*[14] This technology could assist in enhancing transparency and understanding in the decision-making process.[14,15] This transparency is crucial for integrating AI into medical and nursing training, as it can demonstrate how AI models reach diagnoses based on sound medical reasoning. This ensures that future healthcare professionals have the necessary skills to navigate this evolving landscape. Clinicians can evaluate whether the AI algorithm focuses on clinically relevant features or detects spurious patterns, adding transparency and trust, and allowing them to integrate AI into their practice confidently. One beneficial approach is to develop explainable AI (XAI), which involves designing AI systems that can provide accurate diagnoses and explain how they reach those conclusions. This system could even compare those abnormalities to a database of known cases, providing clinicians with evidence-based support for its findings. The main benefit of XAI algorithms in cardiology is the increased trust they foster among clinicians and patients. When clinicians understand how an AI algorithm arrives at a diagnosis, they are more likely to trust its results and use them to inform their decisions. This trust is essential for the widespread adoption of AI in healthcare, and ensures it is used ethically and responsibly.

However, only real collaboration between researchers, clinicians, ethicists, and policymakers can ensure that AI is developed and implemented to genuinely benefit patients and the healthcare system. The acceptance of AI in healthcare hinges on trust in the AI algorithms utilised. Trust can be established through the transparency and interpretability of the models employed.[18]

Conclusion

The primary and most apparent limitation is the AI 'black box' problem, where the inner workings of AI algorithms lack transparency. This lack of transparency poses challenges within the clinical practice setting and could be considered an ethical barrier to implementation. Applying Grad-CAM with selected AI algorithms can significantly bridge this gap. This technique can help clinicians understand how AI algorithms detect or predict AF, enabling them to address patient questions about AI-driven decisions. Consequently, policies and guidelines may need to be drafted or revised to maintain high-quality patient care. Stakeholder involvement may also be necessary to secure funding or investment, ensuring algorithm transparency and justification are not compromised

Key messages

- The 'black box' problem in artificial intelligence (AI) poses ethical concerns and reduces clinician confidence in AI-derived diagnoses due to a lack of transparency in decision-making
- Techniques like Grad-CAM (gradient-weighted class activation mapping) can enhance AI transparency by visualising key electrocardiogram (ECG) sections for atrial fibrillation (AF) prediction. This can help clinicians understand the rationale behind AI decisions and facilitate discussions with patients about their diagnoses

Conflicts of interest

None declared.

Funding

None.

Acknowledgement

The researchers thank Mr Matthew Farrugia, a master's degree holder in artificial intelligence, for reviewing the article and providing valuable technical insights.

References

1. 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**:4–20. https://doi.org/10.1161/CIRCRESAHA.120.316340

2. Attia ZI, Noseworthy PA, Lopez-Jimenez F *et al*. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;**394**:861–7. https://doi.org/10.1016/S0140-6736(19)31721-0

3. Noseworthy PA, Attia ZI, Behnken EM *et al*. Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. *Lancet* 2022;**400**:1206–12. https://doi.org/10.1016/S0140-6736(22)01637-3

4. Hygrell T, Viberg F, Dahlberg E *et al*. An artificial intelligence-based model for prediction of atrial fibrillation from single-lead sinus rhythm electrocardiograms facilitating screening. *Europace* 2023;**25**:1332–8. https://doi.org/10.1093/europace/euad036

5. Karatzia L, Aung N, Aksentijevic D. Artificial intelligence in cardiology: hope for the future and power for the present. *Front Cardiovasc Med* 2022;**9**:945726. https://doi.org/10.3389/fcvm.2022.945726

6. Uittenbogaart SB, Verbiest-van Gurp N, Lucassen WA *et al*. Opportunistic screening versus usual care for detection of atrial fibrillation in primary care: cluster randomised controlled trial. *BMJ* 2020;**370**:m3208. https://doi.org/10.1136/bmj.m3208

7. Matsumoto T, Ehara S, Walston SL, Mitsuyama Y, Miki Y, Ueda D. Artificial intelligence-based detection of atrial fibrillation from chest radiographs. *Eur Radiol* 2022;**32**:5890–7. https://doi.org/10.1007/s00330-022-08752-0

8. Kim JY, Kim KG, Tae Y *et al*. An artificial intelligence algorithm with 24-h Holter monitoring for the identification of occult atrial fibrillation during sinus rhythm. *Front Cardiovasc Med* 2022;**9**:906780. https://doi.org/10.3389/fcvm.2022.906780

9. 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. https://doi.org/10.1016/j.compbiomed.2021.105168

10. Bhardwaj A, Budaraju D, Venkatesh P *et al*. A holistic overview of artificial intelligence in detection, classification and prediction of atrial fibrillation using electrocardiogram: a systematic review and meta-analysis. *Arch Comput Methods Eng* 2023;**30**:4063–79. https://doi.org/10.1007/s11831-023-09935-8

11. Raghunath A, Nguyen DD, Schram M *et al*. Artificial intelligence-enabled mobile electrocardiograms for event prediction in paroxysmal atrial fibrillation. *Cardiovasc Digit Health J* 2023;**4**:21–8. https://doi.org/10.1016/j.cvdhj.2023.01.002

12. 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–S107. https://doi.org/10.3349/ymj.2022.63.S93

13. Chen E, Jiang J, Su R *et al*. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. *Heart Rhythm* 2020;**17**:847–53. https://doi.org/10.1016/j.hrthm.2020.01.034

14. Kim Y, Joo G, Jeon B-K *et al*. Clinical applicability of an artificial intelligence prediction algorithm for early prediction of non-persistent atrial fibrillation. *Front Cardiovasc Med* 2023;**10**:1168054. https://doi.org/10.3389/fcvm.2023.1168054

15. Jo Y-Y, Cho Y, Lee SY *et al*. Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *Int J Cardiol* 2021;**328**:104–10. https://doi.org/10.1016/j.ijcard.2020.11.053

16. Kaminski AE, Albus ML, Ball CT *et al*. Evaluating atrial fibrillation artificial intelligence for the ED: statistical and clinical implications. *Am J Emerg Med* 2022;**57**:98–102. https://doi.org/10.1016/j.ajem.2022.04.032

17. Muralidharan A. AI and the need for justification to the patient. *The Journal of Hospital Ethics* 2023;**9**:55. Available from: https://www.medstarhealth.org/-/media/project/mho/medstar/pdf/mwhc-journal-of-hospital-ethics/johe_v9n2.pdf

18. Stacy J, Kim R, Barrett C *et al*. Qualitative evaluation of an artificial intelligence-based clinical decision support system to guide rhythm management of atrial fibrillation: survey study. *JMIR Form Res* 2022;**6**:e36443. https://doi.org/10.2196/36443