Bridging the Gap Between Artificial Intelligence Research and Clinical Practice in Cardiovascular Science: What the Clinician Needs to Know

Emily Shipley [1,2] Martha Joddrell [1,2] Gregory YH Lip [1,2] and Yalin Zheng [1,3]

1. Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart and Chest Hospital, Liverpool, UK;
2. Department of Cardiovascular and Metabolic Medicine, University of Liverpool, Liverpool, UK;
3. Department of Eye and Vision Science, University of Liverpool, Liverpool, UK

Artificial intelligence (AI) and its applications in cardiovascular science have rapidly grown in recent years, with 77% more papers being published on the PubMed database in 2021 than in 2011 including the terms AI or machine learning (ML) and cardiovascular disease. An even steeper increase can be seen in papers including AI/ML and arrhythmias, with 89% more papers published in 2021 than in 2011.

Advancements have been aided by new ML techniques and increasing computing powers in the form of graphics processing units (GPUs), and availability of large databases, such as the UK Biobank.[1]

Increased implementation of these techniques in clinical practice would have the potential to significantly improve the management of arrhythmias for patients and clinicians alike, but is subject to significant obstacles. *Figure 1* gives an overview of these benefits and concerns associated with AI in clinical practice.

Case for Artificial Intelligence in Arrhythmia Management
Patient Outcomes and Workload Reduction
Disease detection is a key area where AI could prove beneficial. The application of ML to ECG analysis provides a promising solution to lessen the demand placed on clinicians' already-limited time. For AF, early detection and intervention are key to minimising the increasing risk of adverse outcomes and the healthcare costs associated with this arrhythmia.[2]

The use of AI can vastly reduce a medical practitioner's workload as well as improve prognosis by early identification, diagnosis and appropriate management. For example, neural networks can be trained to analyse ECGs, and reinforcement learning can be used to help make dosage decisions.[3] This is important given the move towards a more holistic or integrated approach to AF care, for which studies have demonstrated an association with improved clinical outcomes.[4]

AI-based risk prediction models can identify and quantify risk factors with higher accuracy than traditional risk scores and enable the detection of factors that researchers are unaware correlate with an outcome. For example, logistic regression, gradient boosting, a decision tree and a neural network for stroke risk prediction obtained areas under the curve (AUCs) of 0.891, 0.881, 0.881 and 0.859, respectively, compared with a value of 0.780 obtained by the CHA_2DS_2VASc score.[5]

More widespread use has the potential to improve patient-centred care by further individualising a patient's level of risk, thus enabling the management of modifiable risk factors. An added benefit would be the ability to account for the dynamic nature of risk in certain cardiovascular outcomes. For example, ML and the use of mobile health data could enable stroke risk prediction to adapt to treatment changes over time and incident risk factors, in contrast with the static nature of current standard risk scores.[5]

Data-driven Performance
AI applications are benefiting from the explosion of data creation happening currently. With new methods of collection, data are becoming more diverse, enabling improvement in performance of ML models.

Mobile health data applications have already aided prediction of arrhythmias including AF and VF, as well as supraventricular ectopic beat and ventricular ectopic beat.[6]

ML techniques such as natural language processing (NLP) have enabled researchers to make better use of the data in patients' electronic health records. A study of >63 million individuals applied NLP to free-text data

Figure 1: Benefits and Concerns of Artificial Intelligence in Cardiovascular Research
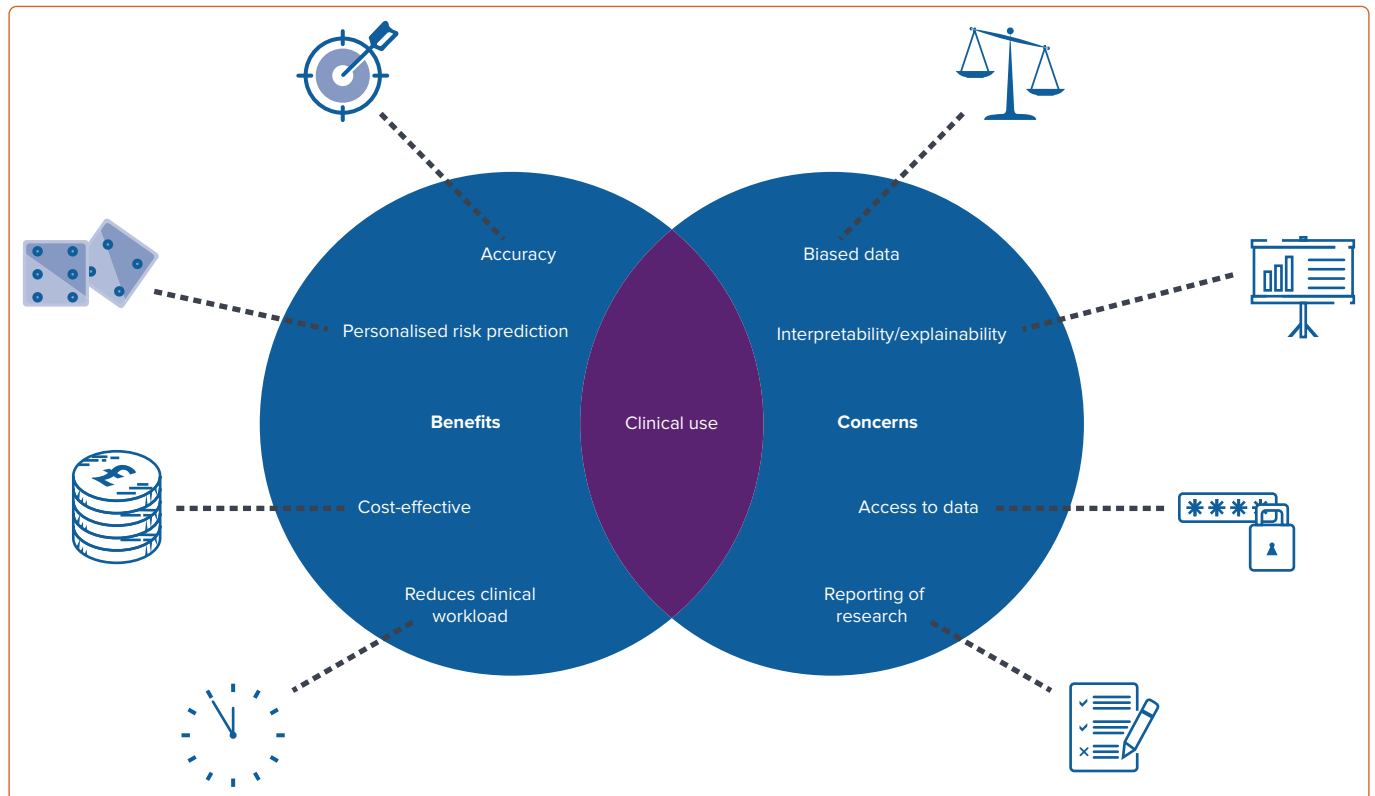
Potential successes that artificial intelligence could have in cardiovascular research with corresponding concerns of use in clinical practice.

combined with structured electronic health record data, and correctly detected 3,976,056 further non-valvular AF cases, compared with using structured data alone.[7]

Evidently, introducing AI-based detection methods into clinical use could help clinicians screen a vast number of arrhythmia cases that may otherwise have gone undetected and, with appropriate treatment, reduce the likelihood of adverse outcomes in these patients.

Barriers to Clinical Use and Potential Solutions
Clinical Acceptance
As AI progresses, such tools are beginning to become accepted into clinical use. For example, in 2019, HeartFlow received approval from the Food and Drug Administration (FDA) to implement its non-invasive, real-time, virtual modelling tool for coronary artery disease intervention.[8]

Although the application of this tool has been a success, agencies such as the FDA can often limit the progress of AI systems by subjecting them to lengthy acceptance processes. The *Artificial Intelligence/Machine Learning (AI/ML) Software as a Medical Device (SaMD) Action Plan* proposes a regulatory framework for the use of AI and ML solutions in healthcare.[9] Better, more explicit resources on approval processes, such as this regulatory framework, would enable researchers and developers to reduce delays and rejection by ensuring their applications meet approval requirements.

Another barrier to the real-world application of AI research is the standard of reporting. Reviews have observed that studies developing prediction models for clinical use are not providing transparent pictures of their methods and findings.[10] As a result, these findings may not be trusted by patients or clinicians, used or replicated.

To address this, laws and protocols are being developed to advise AI researchers on how to thoroughly present their work, such as TRIPOD-AI and PROBAST-AI.[11] They provide guidelines and tools that analysts should follow to prevent research waste and help readers identify key information to make a clear decision on the quality of the studies.

Research waste can also occur when applications are not designed with the needs of clinical purpose at the forefront. Regardless of outstanding performance in predicting an outcome, a model will not be deployed if a clinician requires the prediction of multiple outcomes simultaneously.

Shortcomings of Machine Learning
The complexity of the relationships modelled by deep learning, while a benefit of the technique, may also preclude its use in clinical practice. Applicability decreases if a model generalises too strongly to its training data, resulting in reduced performance with differing populations.

This may be overcome by creating training models using a variety of datasets, although problems arise when health datasets differ in features, even when the same variables are collected. For example, if a model is trained on a dataset that categorises alcohol consumption by <5 units/week and ≥5 units/week, retraining with a dataset that categorises alcohol consumption by <2 units/week, 2–7 units/week, and ≥7 units/week would prove difficult. Feature selection methods, as well as model calibration, provide a solution to this by preventing overfitting.

Similarly, any patterns, biases or outdated information in the data will influence a model's robustness.

Incorrect and harmful applications could ensue if models are applied to unsuitable populations, which may be especially concerning in

cardiovascular science given the life-and-death nature of many clinical decisions.

Gaining access to reflective, diverse data is another problem within AI because of the red tape of data privacy and protection legislation. Developing countries and isolated communities tend to have a lower rate of data collection hence are less likely to be incorporated into model training.[12] As a result, individual personal complexities that could influence outcomes in these populations will not be reflected by models.

Additionally, algorithms built on a specific type of patient data gained from a congruent sample may not apply to those who differ in predictor values. This caveat is being slowly alleviated, with the growth of big data producing more representative and thorough datasets.

These issues, if not addressed, will present significant obstacles to the many potential benefits the clinical use of deep learning could bring.

Interpretability and Explainability

The increasing concern surrounding interpretability and explainability is potentially the most considerable barrier to acceptance of AI in healthcare.

Interpretability refers to the ability to observe the cause-and-effect relationships a model has learned and the outcomes that various factors will produce, e.g. a model predicts that a patient will develop lung cancer as the patient is a current smoker.

Explainability refers to how well the influence of a model's parameters on its decision can be understood; a regression model's coefficients explain that smoking results in a certain increase in the likelihood a patient has lung cancer, whereas a 'black box' deep neural network's weights give a much less explicit insight of how the presence of smoking impacted the final prediction.

For risk prediction in particular, explainability can enable clinicians and patients to mitigate risk by identifying and managing the risk factors contributing most to the prediction.

Increased interpretability and explainability may also help highlight any biases embedded within the data by allowing the examination of a model's choices. Various methods are being developed to help explain the decisions made by more complex models.

For risk prediction, Shapley values can be used to quantify each variable's contribution.[13] However, their calculation is computationally expensive, and the computation time required may be too expensive for a clinician to accept. For deep neural networks, saliency maps can be used to produce visualisations that highlight the patterns and areas of each beat that contribute to the model's final outcome prediction.[14]

Conversely, some researchers question the necessity of interpretability, and argue that the pursuit of interpretable models is not necessary and holding back progression.[15]

There is a strong argument for interpretability for models making high-stakes decisions such as treatment recommendations and dosage calculations, where the negative consequences of an incorrect decision may be substantial. However, for models designed to perform smaller tasks, such as the annotation of ECGs, interpretability may not be essential as long as the model demonstrates good performance.

Responsibility for decisions made by uninterpretable models is a discussion starting to arise now that system manufacturers are seeking approval for them to be deployed in a clinical setting. Since legislation on data and automatic systems are only a recent issue, there are no clear guidelines as to who would be at fault if a misdiagnosis were made by the machine. Although some believe this is a reason to delay use of algorithms in practice, a perquisite is that it takes blame away from overworked medical personnel who, by human nature, are bound to make mistakes occasionally.

Conclusion

Despite its barriers, one cannot deny the success, elevated accuracy and promise that AI is yielding in arrhythmia research. For this progress to be most useful in healthcare, it is imperative that the wall between AI research and clinical care be broken down, through both the implementation of solutions discussed here and the innovation of solutions. ❑

1. Elliott AD, Linz D, Mishima R, et al. Association between physical activity and risk of incident arrhythmias in 402 406 individuals: evidence from the UK Biobank cohort. Eur Heart J 2020;41:1479–86. https://doi.org/10.1093/eurheartj/ehz897; PMID: 31951255.
2. Burdett P, Lip GYH. Atrial fibrillation in the United Kingdom: predicting costs of an emerging epidemic recognising and forecasting the cost drivers of atrial fibrillation-related costs. Eur Heart J Qual Care Clin Outcomes 2022;8:187–94. https://doi.org/10.1093/ehjqcco/qcaa093; PMID: 33346822.
3. Olier I, Ortega-Martorell S, Pieroni M, Lip GYH. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc Res 2021;117:1700–17. https://doi.org/10.1093/cvr/cvab169; PMID: 33982064.
4. Romiti GF, Pastori D, Rivera-Caravaca JM, et al. Adherence to the 'atrial fibrillation better care' pathway in patients with atrial fibrillation: impact on clinical outcomes – a systematic review and meta-analysis of 285,000 patients. Thromb Haemost 2021;122:406–14. https://doi.org/10.1055/a-1515-9630; PMID: 34020488.
5. Lip GYH, Tran G, Genaidy A, et al. Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. Eur Heart J Qual Care Clin Outcomes 2021. https://doi.org/10.1093/ehjqcco/qcab037; PMID: 33999139; epub ahead of press.
6. Sadrawi M, Lin CH, Lin YT, et al. Arrhythmia evaluation in wearable ECG devices. Sensors (Basel) 2017;17:2445. https://doi.org/10.3390/s17112445; PMID: 29068369.
7. Elkin PL, Mullin S, Mardekian J, et al. Using artificial intelligence with natural language processing to combine electronic health record's structured and free text data to identify nonvalvular atrial fibrillation to decrease strokes and death: evaluation and case-control study. J Med Internet Res 2021;23:e28946. https://doi.org/10.2196/28946; PMID: 34751659.
8. Qayyum S, Habib A, Kechyn S, et al. P01 Heartflow: experience of a high-volume district general hospital. Heart 2020;106(Suppl 3):A5.
9. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML) Software as a Medical Device (SaMD) Action Plan. Silver Spring, MD: FDA, 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed 15 March 2022).
10. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011;9:103. https://doi.org/10.1186/1741-7015-9-103; PMID: 21902820.
11. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008. https://doi.org/10.1136/bmjopen-2020-048008; PMID: 34244270.
12. Yan Y, Zhang JW, Zang GY, Pu J. The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine? J Geriatr Cardiol 2019;16:585–91. https://doi.org/10.11909/j.issn.1671-5411.2019.08.010; PMID: 31555325.
13. Wang S, Li J, Sun L, et al. Application of machine learning to predict the occurrence of arrhythmia after acute myocardial infarction. BMC Med Inform Decis Mak 2021;21:301. https://doi.org/10.1186/s12911-021-01667-8; PMID: 34724938.
14. Vijayarangan S, Murugesan B, Vignesh R, et al. Interpreting deep neural networks for single-lead ECG arrhythmia classification. Annu Int Conf IEEE Eng Med Biol Soc 2020:300–3. https://doi.org/10.1109/EMBC44109.2020.9176396; PMID: 33017988.
15. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 2019;49:15–21. https://doi.org/10.1002/hast.973; PMID: 30790315.