A Counterblast to Pessimists and Naysayers – Intelligent Echocardiography Remains the Foundation Stone of Evidence-Based Clinical Cardiology

Alan G Fraser

Abstract

Provocative comments can entertain and instruct as long as they are used to stimulate a civilized discussion, and it is fun to embrace an opportunity to change one's mind (and learn). I am therefore delighted to respond to Adrian Ionescu's comments, although I think he has got it wrong—as I will aim to demonstrate. In the spirit of this debate, please indulge me while I too let off some steam!

I have always disliked the fact that one of the subspecialties within cardiology, which did not exist when I qualified in the 1970s, has come to be known as "cardiac imaging." Cardiac diagnosis is not about pictures, although some conditions are indeed instantly recognizable. Usually, what we need to know to understand disease is how the heart is functioning, much more than what it looks like. That is true for coronary arteriography as much as for non-invasive imaging. If I am forced to adopt a subspeciality label, then I would much prefer to be considered a clinical pathophysiologist.

Accurate diagnosis is the sine *qua non* of logical evidence-based clinical practice, yet we often get it wrong. And there remain many patients with disease that we cannot diagnose precisely because we do not understand it sufficiently. Why does this patient with heart failure with reduced ejection fraction have impaired left ventricular function? Why does that patient with normal blood pressure have left ventricular hypertrophy? In this patient in sinus rhythm, which particular aspects of cardiovascular function will influence the development of dementia? Cardiologists who are expert in performing, analyzing, and interpreting detailed echocardiographic and cardiovascular investigations are needed to give us the best chance of answering such questions. They cannot be replaced by an uninterpretable computer algorithm when no-one yet knows the answer—but by staying in control, researchers can use artificial intelligence (AI) to help their thinking.

Combatting misconceptions about artificial intelligence

1. It is artificial, but it is not intelligent.

All diagnostic imaging systems now generate a huge quantity of digital data from each study. Images are processed and post-processed to improve their quality, using software that is specific to each company because it has been developed by its own engineers. Algorithms are protected as intellectual property, so they are not shared with other companies or made available to clinicians.

Nowadays, a diagnostic study in an individual patient usually involves collecting imaging data in three dimensions over time, with the goal not of making simple single measurements but of analyzing many aspects of function, perhaps both at rest and during stress. The challenges of coping with this mass of data are compounded when images are collected and combined, perhaps from many patients. It is entirely understandable that we need new methods to make sense of such vast imaging biobanks.

There is a strong argument that AI consists of advanced statistical methodologies[1] that have had to be developed because previous methods can no longer cope with the avalanche of data. That concept is much more prosaic than the current fashion and hype which brand AI as a paradigm shift in technology. The vocabulary that is now used to describe AI is misleading and anthropomorphic, since the plain truth is that no computer can "understand" what is does.[2] A computer is not a conscious organism but a machine whose sole function is to perform the tasks that it has been programmed to perform. If that task happens to be running an AI algorithm, it still consists only of instructions to process bits (of binary code) and to perform FLOPS (floating point operations per second) according to instructions prepared by a human being. If an algorithm can "self-learn," it is only because it has been programmed to do that. The real intelligence is always human, not artificial.

2. Generalizations about AI are misleading.

It is misguided to refer to AI as if it is a single phenomenon, since the term encompasses anything from "general" AI (referring to a program intended to mimic human performance by generating "thoughts" and replicating actions) to machine learning (ML) algorithms to deep neural networks. For applications in medicine, the definition of the World Health Organisation may be most useful, namely "the ability of algorithms to learn from data so that they

*Corresponding author: Alan G Fraser, Consultant Cardiologist, University Hospital of Wales, and Emeritus Professor of Cardiology, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XW, U.K.; E-mail address: fraserag@cf.ac.uk; ORCID: 0000-0001-7083-6995

 https://doi.org/10.2478/rjc-2023-0024

can perform automated tasks."[3] Note the absence of the word intelligence from this description.

Smart cardiologists should avoid talking about AI in broad terms, without qualification.

Furthermore, using AI is not a primary objective, whereas using a specific type of AI when it is the best method of performing a particular task or answering a particular question is laudable, perhaps even essential. It is wiser to use a precise term for each specific application.

3. General AI is not a threat to clinical practice.

Recent public concern and debate around the world has been driven by the rapid development of "generative" AI tools for language processing (such as the "Chat Generative Pre-trained Transformer," or ChatGPT, in particular) with their perceived and unknown risks—but as far as we know, they are not used by diagnostic imaging systems. In addition, true general AI does not yet exist, and it seems very unlikely that it would ever be implemented into some aspect of clinical practice because of the implications of ceding human control.

Software including AI algorithms with a medical function, such as making a diagnosis, planning treatment, or predicting outcomes, is designated in the European Union as a medical device (SaMD, or "software as a medical device").[4] Whether it is a standalone program or it is incorporated within an implantable device, it has to be approved by a notified body and given a CE mark (for Conformité Européenne) before it can be implemented in practice. High-risk SaMD that is used in critical clinical situations, when any erroneous operation could lead to potentially serious consequences for the patient, will only be approved if it has undergone a satisfactory clinical evaluation with evidence of substantial benefit that outweighs any risks. All major jurisdictions for medical devices are developing regulations to ensure that high-risk AI or ML must meet certain minimum standards.[5] The engagement and collaboration of expert clinicians in supporting regulatory decisions can ensure that medical ML algorithms are approved only when they are needed and will benefit patients. The European Society of Cardiology is leading a European consortium which has amongst its tasks a project to define the criteria on which such decisions should be based.[6,7]

Exploiting the clinical potential of AI

The vast majority of AI medical devices which have been approved by regulators are software applications for diagnostic imaging that are already in routine use for tasks such as optimizing image quality, detecting and labeling structures (called "segmentation" by AI engineers), and performing measurements. Systematic reviews suggest that ML imaging tools can outperform novices and attain levels of accuracy that are equivalent to experts, but more widely there is no convincing evidence yet that AI-based decision support improves clinical outcomes.[8] The development of ML could indeed have a "devastating impact" on clinical expertise if it is left to engineers rather than being viewed as an exciting challenge and opportunity for collaborating to improve the quality of diagnostic practice. The priority must be to ensure that new software tools are developed, validated, and implemented appropriately—which means that diagnostic imaging experts will continue to be needed, both to set the agenda and to introduce new tools safely.

1. Clinical imperatives for safe implementation

The first essential requirement is transparency.[9] Whenever an AI tool or ML algorithm is being applied, its use must be announced. It should be easy to discover details of the training dataset that was used for development and whether the tool was ever retested in an independent population. Diagnostic imaging systems should indicate to the clinician if the ML tool has been sufficiently validated to be applied in each individual case.

The precision and reproducibility of ML software is frequently exaggerated by comparing repeated measurements of the same dataset for each subject, instead of analyzing two separately acquired sets of images of the same structures. The result of using ML in any individual patient should therefore be reported with its diagnostic error or confidence intervals. Measurements should not be oversimplified (such as reporting a single number for ejection fraction without giving left ventricular volumes, when it is a ratio with limited utility). Recommendations from AI systems will only be accepted in clinical practice if they are intelligible; extreme caution is required when considering a diagnosis ("classification") offered without explanation by a neural network.

2. Using AI tools in medical research

There has been a tendency during recent decades to fund experimental more than clinical research, but early hopes of major advances through better understanding of basic mechanisms of disease have not all been fulfilled. The promise of widespread health benefits from the genomic revolution, for example, has yet to be fulfilled. It has become recognized that disease may result from the net effects of complex interactions affecting common final pathways, with genetic influences, although significant, accounting for only a small proportion of the prevalence of common (rather than monogenic) diseases.[10] For further advances, patients need to be characterized by precise clinical phenotyping and their images interpreted using profound knowledge of clinical pathophysiology.

When a disease is poorly understood, then unsupervised machine learning may identify phenogroups of patients who share features that cannot be recognized using conventional imaging and statistical techniques, but it is important that the results are interpretable. Otherwise, the underlying pathophysiological mechanisms may remain obscure, and it will be impossible to develop and test treatments that will target the key steps.

An example could be heart failure with preserved ejection fraction, for which most clinical trials are widely held to have failed despite them having been performed in subjects who were recruited using suboptimal diagnostic entry criteria, such as left ventricular ejection fraction and dyspnoea. Initial studies with ML have only reinforced what we already knew.[11] Advances in treatment will remain elusive until trials are more focused, which will be possible only with the participation of clinical diagnostic specialists. AI and ML tools may

provide new insights and augment their knowledge, but they will not replace their expertise.

"Cardiac imaging" as a career?

Every cardiologist—not none—should be an "imager", in the sense that every clinician who treats patients needs to understand the mechanisms of their symptoms and pathology. Sub-specialization will have gone too far if it becomes widely accepted that the insights obtainable from detailed functional imaging performed by an experienced echocardiologist can be replaced by an automated report from a machine that has been generated by ML. That would also be an indictment of the suboptimal quality of many current diagnostic studies.

It is wrong to argue in favour of one diagnostic imaging modality rather than another—for example, for echocardiography instead of computed tomography (CT) or magnetic resonance (MR) imaging. But it is equally wrong to imply that CT or MR are better because they produce nicer tomographic images that can be obtained with less dependence on practical skills. The array of options is complementary, but CT and MR cannot replace the detailed physiological insights that echocardiography is capable of providing with precision at high frame rates. Echocardiography has reached a state of maturity, but its utility is still not fully exploited. Don't abandon it—exploit it!

The need for (truly) intelligent cardiologists will never be replaced by (falsely) "intelligent" machines. And patients want to be cared for by fellow human beings with whom they can talk and whom they can trust. If we hand control of echocardiography to computer algorithms, we might as well abolish doctors altogether. Patients could be diagnosed by machines and algorithms, and treated according to guidelines or undergo procedures by robots—is that really what we want?

Conflict of interest

none declared.

REFERENCES

1. Faes L, Sim DA, van Smeden M, Held U, Bossuyt PM, Bachmann LM. Artificial intelligence and statistics: just the old wine in new wineskins? Front Digit Health. 2022; 4: 833912.
2. Bishop JM. Artificial intelligence is stupid and causal reasoning will not fix it. Front Psychol. 2021; 11: 2603. www.frontiersin.org/article/10.3389/fpsyg.2020.513474
3. World Health Organization. Ethics and governance of artificial intelligence for health. WHO Guidance, Geneva 2021. https://www.who.int/publications/i/item/9789240037403
4. European Commission, Medical Device Coordination Group. MDCG 2019-11 Guidance on qualification and classification of software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. October 2019. https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_software_en_0.pdf
5. Fraser AG, Biasin E, Bijnens B, Bruining N, Caiani EG, Cobbaert K, Davies RH, Gilbert SH, Hovestadt L, Kamenjasevic E, Kwade Z, McGauran G, O'Connor G, Vasey B, Rademakers FE. Artificial intelligence in medical device software and high-risk medical devices - A review of definitions, expert recommendations and regulatory initiatives. Expert Rev Med Devices. 2023; 20: 467–491.
6. Coordinating Research and Evidence for Medical Devices (CORE-MD) is an EU Horizon 2020 project (965246). For more information visit www.core-md.eu
7. Fraser AG, Nelissen RGHH, Kjærsgaard-Andersen P, Szymański P, Melvin T, Piscoi P; CORE-MD Investigators. Improved clinical investigation and evaluation of high-risk medical devices: the rationale and objectives of CORE-MD (Coordinating Research and Evidence for Medical Devices). Eur Heart J Qual Care Clin Outcomes. 2022; 8(3): 249–258.
8. Vasey B, Ursprung S, Beddoe B, Taylor EH, Marlow N, Bilbro N, Watkinson P, McCulloch P. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. JAMA Netw Open. 2021; 4(3): e211276.
9. Fraser AG, Butchart EG, Szymański P, Caiani EG, Crosby S, Kearney P, Van de Werf F. The need for transparency of clinical evidence for medical devices in Europe. Lancet. 2018; 392: 521–530.
10. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, Kathiresan S, Kenny EE, Lindgren CM, MacArthur DG, North KN, Plon SE, Rehm HL, Risch N, Rotimi CN, Shendure J, Soranzo N, McCarthy MI. A brief history of human disease genetics. Nature. 2020; 577: 179–189.
11. Fraser AG, Tschöpe C, de Boer RA. Diagnostic recommendations and phenotyping for heart failure with preserved ejection fraction: knowing more and understanding less? Eur J Heart Fail. 2021; 23: 964–972.