Artifical Intelligence and Digital Innovations in Cardiovascular Care 1

Challenges for augmenting intelligence in cardiac imaging

Partho P Sengupta, Damini Dey, Rhodri H Davies, Nicolas Duchateau, Naveena Yanamala

Artificial Intelligence (AI), through deep learning, has brought automation and predictive capabilities to cardiac imaging. However, despite considerable investment, tangible health-care cost reductions remain unproven. Although AI holds promise, there has been insufficient time for both methodological development and prospective clinical trials to establish its advantage over human interpretations in terms of its effect on patient outcomes. Challenges such as data scarcity, privacy issues, and ethical concerns impede optimal AI training. Furthermore, the absence of a unified model for the complex structure and function of the heart and evolving domain knowledge can introduce heuristic biases and influence underlying assumptions in model development. Integrating AI into diverse institutional picture archiving and communication systems and devices also presents a clinical hurdle. This hurdle is further compounded by an absence of high-quality labelled data, difficulty sharing data between institutions, and non-uniform and inadequate gold standards for external validations and comparisons of model performance in real-world settings. Nevertheless, there is a strong push in industry and academia for AI solutions in medical imaging. This Series paper reviews key studies and identifies challenges that require a pragmatic change in the approach for using AI for cardiac imaging, whereby AI is viewed as augmented intelligence to complement, not replace, human judgement. The focus should shift from isolated measurements to integrating non-linear and complex data towards identifying disease phenotypes—emphasising pattern recognition where AI excels. Algorithms should enhance imaging reports, enriching patients' understanding, communication between patients and clinicians, and shared decision making. The emergence of professional standards and guidelines is essential to address these developments and ensure the safe and effective integration of AI in cardiac imaging.

Lancet Digit Health 2024; 6: e739–48

Published Online August 29, 2024 https://doi.org/10.1016/S2589-7500(24)00142-0

This is the first in a Series of four papers about artifical intelligence and digital innovations in cardiovascular care. All papers in the Series are available at www.thelancet.com/series/AI-and-digital-tools-in-cardiovascular-care.

Division of Cardiovascular Disease and Hypertension, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA (Prof P P Sengupta MD, N Yanamala PhD)**; Biomedical Imaging Research Institute, Cedars–Sinai Medical Center, Los Angeles, CA, USA** (Prof D Dey PhD)**; Institute of Cardiovascular Science, University College London, London, UK** (R H Davies PhD)**; CREATIS, INSA, CNRS UMR 5220, INSERM U1294, Université Lyon 1, UJM Saint-Etienne, Lyon, France** (N Duchateau PhD)**; Institut Universitaire de France, Paris, France** (N Duchateau)

Correspondence to: Prof Partho P Sengupta, Division of Cardiovascular disease and Hypertension, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA **partho.sengupta@rutgers.edu**

Introduction

Cardiovascular diseases are a leading cause of death globally and the use of cardiac imaging for accurate assessment of cardiac structure and function is crucial for diagnosis, treatment planning, and prognosis. Artificial intelligence (AI) has rapidly permeated the field of cardiac imaging. Its burgeoning influence, however, is not devoid of complexities. In 2022, a National Heart, Lung, and Blood Institute workshop highlighted that even as AI applications in cardiac imaging continue to expand, there is not yet substantial proof that these developments can meaningfully reduce health-care costs.[1] Furthermore, beyond automation, there is not enough evidence from prospective, blinded, randomised clinical trials to illustrate that these technologies are superior to human interpretation or substantially affect patient outcomes. Nevertheless, there is a competitive race within the academic and industrial sectors to create superior technological solutions for cardiovascular disease using AI. In this Series paper, however, we take a more pragmatic approach. Although AI might support clinicians in specific tasks, it cannot replace the nuanced clinical judgement acquired through years of experience. The American Medical Association defines AI as augmented intelligence to reflect its perspective that AI-based tools and services support, rather than explicitly replace, human decision making.[2] Adopting this perspective is crucial for setting realistic expectations and ensuring AI's safe and practical application in cardiac imaging.

Evolution of AI in cardiac imaging

There is strong clinical rationale for the introduction of AI techniques to automate cumbersome measurements in cardiac imaging, and thereby improve workflow efficiency. Regarding the techniques themselves, for imaging data, deep learning supersedes more conventional machine learning as it automatically finds the appropriate features in the images while solving the targeted problem (figure). Deep neural networks are universal function estimators,[3] meaning that they can find the complex relationship (often non-linear) between the input data (eg, image and clinical patient characteristics) and the output (eg, segmented regions and diagnosis). In addition to their predictive performance, neural networks are fast at inference (ie, predicting the output for new cases not used during training) and are robust (ie, they generalise well) if given enough data. Examples include cardiac image segmentation[4] or the approximation of complex differential equations used in biophysical simulations.[5–7]

AI methods are often criticised for being uninterpretable black boxes, which can restrict clinical acceptance. However, the route towards a prediction does not always need to be explicitly interpretable (eg, myocardial segmentation when the wall is visible in the images). Besides, many researchers have now gained an understanding of how specific features of complex deep learning architectures work in cardiac imaging (eg, convolutions, residual blocks[8–11] and, at a broader scale, encoding and decoding mechanisms). The effects of sub-blocks in AI architectures (eg, convolutional or residual blocks) and their combinations (eg, contracting and expanding paths in a U-net to operate at multiple scales) on model performance are now well understood by data scientists. More interpretable solutions are being explored, either within the model itself or for post hoc analysis (eg, Gradient-weighted class activation mapping [Grad-CAM], feature importance scores, Shapley additive explanations, and local interpretable model-agnostic explanations), although some of these methods are currently undergoing development and evaluation.[12]

There is growing interest in using the next generation of AI models inspired by natural language processing, such as vision transformers, which can handle complex structures in the data both for the input and the output, as shown for medical image segmentation and classification.[13] Transformer architecture, one of the most popular technological advances in deep learning,[14,15] is inspired by natural language processing, is able to assimilate a structured group of words to produce realistic texts or images, and is already capable of succeeding in postgraduate cardiology examinations.[16] The generative pre-trained transformer models are at the spearhead of these advances. At the core of these methods are attention mechanisms that weigh the importance of different inputs and reinforcement learning mechanisms, which enable improved learning performance from human-expert feedback.

Key messages

- Deep learning techniques are being increasingly applied to cardiac imaging for automating measurements and improving workflow efficiency
- Despite the growth of artificial intelligence (AI) in cardiac imaging, there is not enough evidence showing its cost-effectiveness, superiority compared with human interpretation, or improvements to patient outcomes
- Challenges include data scarcity, lack of data diversity, evaluation difficulties, misalignment with stakeholders, and issues with regulatory approval, data stewardship, and data privacy; professional standards are emerging to address these constraints
- It is essential to recognise the limitations of AI and understand that, at present, it can support imagers to reduce repetitive low-calibre activities, but not replace their nuanced clinical interpretation
- A framework shift in cardiac imaging should move the focus away from isolated measurements and leveraging AI's pattern recognition capabilities to integrating complex, non-linear data for precise disease phenotype identification
- Future innovative algorithms, such as large language and vision transformer models, could play a pivotal role in enhancing imaging reports, facilitating patient understanding, improving communication between patients and clinicians, and supporting collaborative decision making

Landmark developments in using AI for cardiac imaging

The table presents a numerical estimate of AI and machine learning applications in cardiac imaging that have received regulatory approval and are currently available in the USA, as reported by the US Food and Drug Administration (FDA) in October, 2023.[17]

Echocardiography

Deep learning algorithms have been applied to view recognition, segmentation, and assessment of echocardiographic volumetric measurements, including indices such

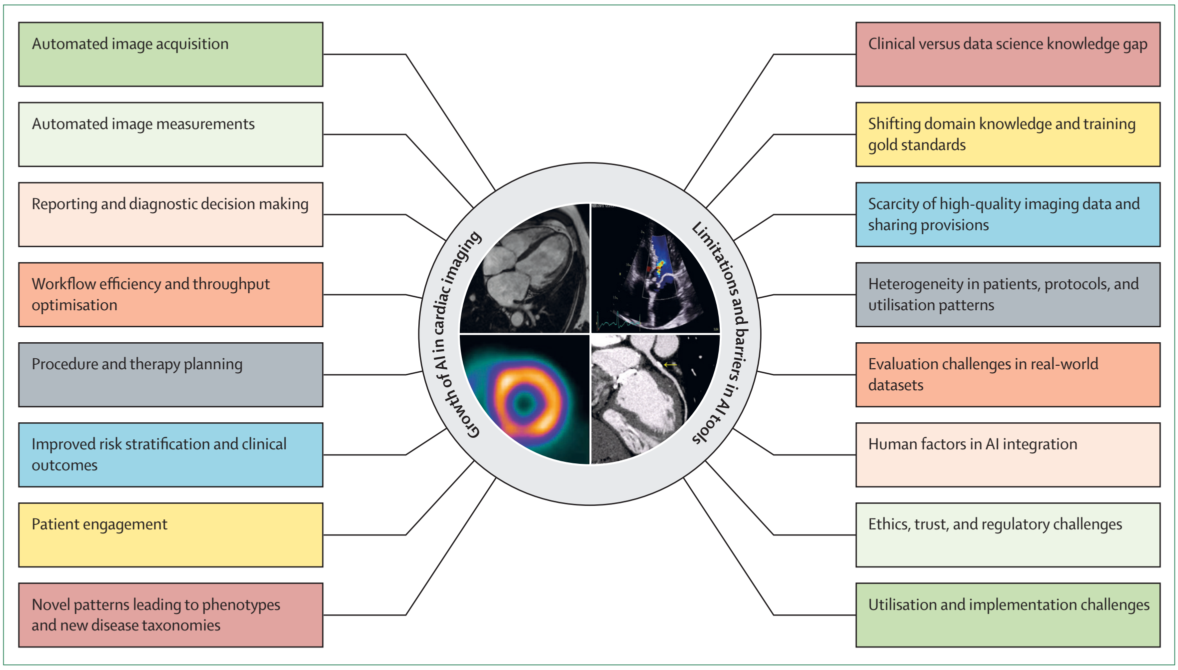

Figure: **Exploring the benefits and challenges for applying AI in cardiac imaging**
AI=artificial intelligence.

as left ventricular ejection fraction (LVEF). A video-based deep learning network has been implemented for fully automated beat-to-beat evaluation of LVEF, even in the presence of arrhythmias.[18] In a blinded, randomised clinical trial of initial AI and sonographer assessment for patients undergoing echocardiographic quantification of cardiac function, initial evaluation of LVEF by AI was non-inferior to assessment by sonographers, with a smaller proportion of the AI group requiring correction for the final cardiologist assessment.[19] AI algorithms have also been applied in resource-constrained settings. In a multicentre study, a deep learning algorithm was developed to guide novice users without experience in cardiac ultrasound to acquire diagnostic-quality transthoracic echocardiographic images.[20] Similarly, another multicentre study introduced a deep learning-based method for quantifying LVEF, focusing on limited views obtained during point-of-care ultrasound imaging.[21]

Although the majority of data are derived from studies of adults, AI-enabled studies have also been reported based on fetal and neonatal cardiac imaging. For example, an ensemble of deep learning algorithms has shown expert-level prenatal detection of complex congenital heart disease from screening echocardiographic images.[22]

Unfortunately, no single echocardiographic parameter adequately captures the complexity of cardiac systolic or diastolic function. As a result, echocardiography interpretation often includes semiquantitative statements regarding complex assessments, which can be subjective and dependent on the expertise of the interpreting clinician. Moreover, AI algorithms might struggle to replicate interpretations from experienced cardiologists who often integrate multiple measurements while discounting any deficits that might result from suboptimal image quality. However, unsupervised clustering algorithms have been applied to reduce heuristic biases related to methods for integrating complex echocardiographic measurements, which improves the classification of patient subgroups who have a similar risk of future adverse events, such as all-cause and cardiac mortality.[23–25] The application of such AI algorithms has been shown to be superior to existing guideline-based classifications in diagnostic and prognostic value, and might potentially enhance patient care.

	Applications (n)	Companies* (n)
Echocardiography	36	13
Cardiac CT	24	15
Cardiac MRI	7	6
Multimodal imaging	3	2

Applications had the purposes of scanning device, reconstruction, image enhancement, viewer, automatic quantification, analysis tools, or intervention planning. Data were extracted through a manual review of the US FDA report and should be viewed as relative estimates within the limits of the provided information.[17] Ten of 71 applications were tagged as cardiovascular-concerned imaging (the other applications represented echocardiographic analysis or biosensor-based monitoring), and 60 of 531 applications were tagged as radiology-concerned cardiac imaging. Applications that were not specific to cardiac imaging (eg, MRI or CT scanning devices, and generic acquisition-related processing) were excluded. FDA=Food and Drug Administration. *Number of companies with approved applications.

Table: **US FDA-approved artificial intelligence applications for cardiac imaging**

Cardiac CT

AI algorithms can improve workflow efficiency in cardiac CT interpretation. For example, a multicentre study observed a 22% reduction in chest CT interpretation times.[26] Deep learning has been used to automate the quantification of cardiac CT-based imaging biomarkers for enhanced cardiovascular outcome prediction. For example, automated AI-enabled coronary artery calcium scoring—a marker of total coronary atherosclerosis and a strong predictor of future cardiovascular events—using both echocardiography-gated cardiac CT and ungated thoracic CT has recently been introduced in clinical practice. Deep learning can extract quantitative biomarkers from epicardial adipose tissue, a metabolically active fat depot for enhanced prediction of adverse cardiovascular events.[27]

Deep learning has been used to assess the severity of coronary artery stenosis on coronary CT angiography, a first-line modality for evaluating chest pain. This method has the potential to reduce inter-reader variability and interpretative error.[28] An international multicentre study revealed excellent agreement of deep learning with expert readers for total plaque volume in intravascular ultrasound images, at a fraction of the time taken by experts (5·6 s *vs* 25·7 min by experts).[29] Deep learning-based complete-plaque volume assessment showed an increased risk of myocardial infarction in the prospective SCOT-HEART multicentre trial.[29] Deep learning has also been applied for evaluating the functional assessment of coronary stenoses from CT angiography with non-invasive CT angiographic-derived fractional flow reserve (CT-FFR), with comparable accuracy and significantly shorter execution times than computational fluid dynamics-based CT-FFR, which can potentially facilitate CT-FFR calculation at a standard workstation at point of care.[30] A 2023 randomised trial has shown that onsite CT-FFR with deep learning reduced the trial-specified primary endpoint (proportion of patients undergoing invasive coronary angiography without obstructive coronary artery disease and patients undergoing invasive coronary angiography with obstructive coronary artery disease who did not undergo early intervention) when compared with standard of care; however, there was no significant difference in 1-year adverse cardiovascular events.[31]

Cardiac CT has been increasingly used for diagnosis and pre-procedural planning in structural heart interventions. In a multicentre setting, researchers have investigated the use of computational modelling to optimise planning for transcatheter structural interventions, such as left atrial appendage closure[32] and transcatheter aortic valve therapy.[33] The cardiac CT-based computer simulations used in these studies improved procedural outcomes by providing enhanced insights on the potential risks associated with challenging autonomies.

The greatest clinical impact of AI application in cardiac CT imaging has been for segmentation and quantification. But despite reported success with a marked reduction of time required for processing images, variations in images due to differences in acquisition protocols, image quality, and heterogeneity in patient anatomy require the final step to be approved by readers with advanced training in cardiac CT. Moreover, biological variations in diseases result in challenges. For example, atherosclerotic cardiovascular disease presents differently in men and women.[34] Similarly, in addition to absolute measurements of coronary plaque, age-based risk thresholds for coronary calcium concentrations are not fully incorporated into AI models, but doing so could improve personalised risk stratification.

MRI

One of the biggest impacts of AI in cardiac imaging has been in automating image segmentation in cardiac MRI, particularly for the measurement of LVEF.[35] Similar algorithms have enabled large-scale automated analysis of population cohorts such as the UK Biobank and the Multi-Ethnic Study of Atherosclerosis, providing new insights into disease progression and healthy ageing.[36] AI has also allowed full automation of other MRI-derived imaging biomarkers, facilitating routine use in clinical practice. For example, myocardial perfusion reserve, measured at a pixel level using AI, was shown to be associated with death and adverse cardiovascular events on follow-up.[37] Despite the enthusiasm for AI-supported clinical cardiovascular MRI, there are few multi-institutional prospective studies across different clinical teams and vendors that show the clinical benefit that such AI applications could potentially accomplish.

Cardiac nuclear imaging

Deep learning algorithms have been applied to large multicentre registries for automated analyses of nuclear myocardial perfusion imaging to predict substantial coronary artery disease and prognostic outcomes. For example, in a registry of more than 20 000 patients undergoing myocardial perfusion imaging by single photon emission CT (REFINE–SPECT), an AI-based model was shown to outperform conventional perfusion imaging for both time-specific and event-specific predictions of adverse cardiovascular outcomes, with a display of explainable AI probability that could help to identify and modify individual risk factors.[38] For PET myocardial perfusion imaging, an AI model trained directly on myocardial polar maps has been shown to improve patient risk stratification for all-cause mortality in comparison with the current clinical standard for PET flow or perfusion assessments, with attention maps displayed on the polar maps to highlight regions of decreased myocardial perfusion.[39] As with MRI, there are no prospective studies and clinical trials that show the benefit of AI-based predictions for clinical diagnosis and patient outcomes.

Limitations of current AI approaches and barriers to clinical implementation

Clinical versus data science knowledge gap

Despite the huge boom in publications and public communication around AI, clinical translation remains scarce. In other words, many AI algorithms and variants are being developed for cardiac imaging, but few are used in clinical practice. Although there has been extensive research in AI for cardiac imaging, integrating these innovations into clinical practice remains a substantial challenge (see figure, right-hand side). A potential reason for this gap could be the lack of cooperation between data scientists and imaging cardiologists, resulting in clinicians who do not understand data science methods and data scientists who do not understand the nature of clinical application. For example, AI research in cardiac imaging has embraced attention maps to help interpret decisions made by deep learning models.[40]

The most popular approach is saliency maps, which identify the input features that most influence a model's decision-making process,[40] but concentrating on a few features of a complex model only permits a narrow understanding of model behaviour and can lead to a focus on the wrong features.[41,42] The input features might not equate to the presence of an abnormality and the relationship between the imaging features that AI and humans use to classify disease does not always translate linearly. Furthermore, saliency maps are often interpreted visually, which is subjective, and complex patterns can be challenging to interpret and even misleading.[43] For example, a 2022 study revealed that Grad-CAM, a popular method for generating saliency maps, underperforms at accurately localising ten specific pathologies on chest x-rays, particularly when these pathologies exhibit multiple instances, are small, or have complex shapes.[44] These observations further indicate that the reliability of saliency maps increase with the model's confidence in its predictions, suggesting that caution should be applied when using these maps as decision aids in clinical settings.[44] Additionally, over-reliance on data sourced from public competitions and challenges,[45] although valuable for benchmarking, might inadequately capture the complex and multifaceted nature of real-world clinical scenarios. Clinicians are crucial for providing insights into real-world complexities that public datasets might overlook, which is essential for improving the relevance and accuracy of AI tools for clinical implementation.

In addition to these substantial issues, the unbridled hype around AI is associated with the excessive use or misuse of certain terms (eg, AI for simple machine learning and explainable for simple ways to represent the data). AI is not only deep learning; it also includes many other methods, particularly knowledge-based algorithms and reinforcement learning, which could be relevant to clinical data. The application of deep learning is not universally suitable for every health-care problem. The inherent level of complexity of a problem might not always justify the adoption of intricate, deep learning architectures. Although deep learning methods often capture the spotlight, simpler machine learning approaches, such as the XGBoost algorithm for tabular data,[46] can often outperform their more complex counterparts with fewer computational resources. Therefore, it is crucial to embrace the versatility of AI and acknowledge the potential of straightforward methods in delivering robust solutions to complex clinical challenges.

To bridge the gap between AI research and clinical imaging, multidisciplinary collaboration between data scientists and clinicians is key and could be fostered through initiatives that embed data scientists in the clinical field. Clinicians could also benefit from AI training and lessons in data science could be incorporated into medical school curricula. Most importantly, AI applications should be targeted at clinical needs and research funding could be directed towards these areas. For example, developing high-level, easy-to-use machine learning libraries written explicitly for medical imaging (eg, Medical Open Network for AI) can speed up the adoption of AI among clinicians and reduce dependence on specialist AI skills.

Shifting domain knowledge

AI models for cardiac imaging are often built without truly understanding the underlying cardiovascular system. Although data are only one facet of the targeted clinical problems, the incorporation of physiological knowledge could be desirable in many applications. In our pursuit of applying AI techniques to comprehend cardiac structure and function, we encountered one key challenge: the absence of a unified model that fully captures the heart's complex structure and function. Until the early 21st century, more than eight models proposed various arrangements of the heart's muscle fibres, with substantial debates surrounding how the helical structures contributed to cardiac deformations.[47] Despite the insights gained from imaging technologies, a comprehensive model that encompasses all aspects of cardiac function remains elusive. Clinicians perceive the heart in different ways depending upon their field of clinical work (invasive catheter-based approach, imaging, clinical trials addressing neurohumoural pathways, etc). Contemporary views range from viewing the heart as a hydrodynamic pressure pump (with measurements supported by invasive catheters), a squeezing chamber (assessed through ejection fraction and related parameters), or a muscle pump (analysed using strain and complex twisting deformation).[48] These diverse perspectives have influenced clinical guideline recommendations, clinical decisions, and the development of AI algorithms. For example, the continued emphasis on ejection fraction in cardiac imaging and guidelines has led AI-based investigations to focus primarily on this metric. However, the drawbacks of ejection fraction in characterising cardiac function and heart failure syndromes are well recognised.[49] Beyond experts arbitrarily defining heart failure as preserved ejection fraction, midrange ejection fraction, and reduced ejection fraction, societal and clinical trial cutoffs for normal LVEF are not uniform. Recent investigations have shown the restrictions of using ejection fraction categories to define individual patient phenotypes.[50] There is a growing recognition that cardiac function or dysfunction are latent or hidden behaviour of the myocardium that isolated parameters cannot measure. Future work should guide AI models with physics-based knowledge (eg, differential equations governing cardiac biophysical models, flow computations, etc) or even physiological models of cardiac function to aid the synergistic integration of scientific knowledge and data into the AI framework.[51]

Furthermore, the heterogeneity of cardiovascular diseases exacerbates AI training, given the fragility of cardiac disease classification and the abundance of unlabelled data. Specifically, the taxonomy of cardiac conditions is limited and has implications for developing accurate disease labels for AI algorithm training data.[24,52] Similar to genetic studies, where the heterogeneity of phenotypic definitions in coronary artery disease has hindered the replication of genetic associations,[53] the challenges of heterogeneity in cardiac disease classification also affects imaging. AI algorithms that aim to interpret health data, including clinical images, genetic datasets, and electronic health records, rely on accurate disease labels to train and optimise performance. However, the variability in disease classification and the absence of standardised definitions pose challenges for AI algorithms in accurately identifying and classifying cardiac disease, and thereby affecting their performance and generalisability. Therefore, it is crucial to address the heterogeneity in cardiac disease taxonomy by establishing more robust definitions open to human variability to ensure the development of accurate disease labels for AI algorithms.

Scarcity of high-quality and standardised data

The availability of curated and high-quality datasets remains the greatest challenge for AI, and furthermore increasing dataset diversity with respect to disease prevalence, racial and gender diversity, comorbidities, and imaging referral patterns and imaging hardware can improve the generalisability of AI-enabled algorithms.[1] Machine learning methods, such as deep learning, perform best when trained on a substantial dataset representing a diverse population. Major AI advances have been made in cardiac cine MRI segmentation, where data have been made freely available for grand challenges[4] or released as part of population cohort studies.[54] However, openly available data remain scarce in many domains and the effort required to collect data is prohibitive, in terms of the high costs and the resource-intensive and labour-intensive nature of the data collection process. The scarcity of clinicians able to annotate training data is a further impedance. Data protection regulations, such as the General Data Protection Regulation in Europe, are positive initiatives that protect patient privacy and personal data security, but they can raise barriers to obtaining data for AI development. Data sharing initiatives, such as the Society for Cardiovascular Magnetic Resonance registry of cardiac MRIs and associated clinical data, could help alleviate these issues. The development of federated and swarm learning can assist institutions in sharing data within these registries while alleviating privacy concerns. However, there is currently little incentive to boost organisational motivation and facilitate sustainable sharing of cardiac imaging data for AI development. Additionally, the clinical value of AI in cardiac imaging could be enhanced by incorporating multimodal data, such as electronic health records data, and adopting initiatives, such as trusted research environments, which could enable data sharing and processing of clinical data within a secure environment.

For more on the **Society for Cardiovascular Magnetic Resonance registry** see https://scmr.org/page/Registry

Obtaining data that sufficiently represent all patients can be challenging, but a specific effort to include under-represented groups is crucial to prevent the perpetuation of bias and improve both performance and generalisability of AI models.[55] Although medical imaging has established standards (eg, Digital Imaging and Communications in Medicine), not all imaging data are standardised and harmonised, and raw data are not permanently stored. Moreover, technological advances such as harmonic imaging in echocardiography and steady-state free procession for cardiac cine MRI have improved data acquisition in cardiac imaging. Although these innovations have improved image quality, the underlying data fundamentally differ from legacy methods, and clinicians interpret and measure them differently.[56] AI interpretation is also likely to be affected by these advances; therefore, developers must train different models for altered acquisition parameters or attempt to harmonise the data.

Evaluation challenges

Evaluating AI algorithms in diverse real-world datasets is crucial to ensure their effectiveness across clinical settings,[57] but obtaining sufficient data before the model's release can be challenging. In such cases, post-market surveillance is essential to monitor the algorithm's performance once used, ensuring that the AI system performs adequately and meets the desired standards.[58] Each AI prediction would ideally be accompanied by its confidence in specific predictions so that this can be factored into clinical decision making, but quantifying uncertainty can be challenging.[59] For example, several factors, including the complexity of the models, variability in real-world data compared with training data, and the difficulty in modelling all potential sources of error or ambiguity, can make it difficult to accurately estimate the confidence levels of predictions.

Choosing appropriate evaluation metrics are essential when evaluating AI for clinical use. Evaluation is often performed by comparing machine learning and deep learning predictions to those made by a clinician. The DICE coefficient, for example, is often used to measure the overlap between AI and clinician image segmentation. However, clinicians' interpretations are subjective and inconsistent, and might not be suitable as gold standards. A shift in focus towards more clinically meaningful evaluation metrics is needed, with the aim of improving clinical outcomes.

The quality of data used for evaluation is dynamic and subject to shifts caused by changes in imaging technology, imaging protocols, or use patterns. These shifts can affect the performance of AI algorithms, highlighting the need for continuous monitoring and adaptation to ensure reliability and effectiveness.[60] Various approaches could be adopted to address these evaluation barriers. Prospective controlled clinical trials and registry studies can be conducted to evaluate the effectiveness of externally validated algorithms in clinical practice and build user trust. These studies can provide insights into the performance and effect of AI algorithms in real-world settings. Additionally, efforts can be made to enhance AI explainability,[61] fostering trust in the algorithms for automated quantification and interpretative tasks.

Human factors

The introduction of an AI system into clinical practice can affect many people. Health-care workers, patients, and the public often misapprehend AI and view it as a hindrance rather than an aid.[62] Substantial investment is needed to engage stakeholders and ensure alignment with the application of AI. AI integration into practice is not just about the computer programme but how it fits into clinical workflows, which might require optimisation of human–system interaction and experience.[63] A human-in-the-loop approach can potentially mitigate risks associated with recommendations made by AI and can help address questions about accountability.[64] This approach requires the system to be interpretable, but this can be difficult to achieve with the complex models of modern AI and can lead to outcomes that are not fully understandable.[65]

Regulatory challenges

Clinical translation of AI tools can be a time-consuming process, sometimes without academic recognition or benefits (eg, publications or career advancement), particularly when obtaining regulatory approval from a certifying body (eg, the US FDA; table) or obtaining a CE mark is needed. Pathways to help traverse this landscape could lower barriers to the clinical translation of AI tools and encourage more academic teams or smaller companies to translate their tools into clinical use. Data stewardship, which involves the responsible management and oversight of data, can also present challenges for AI algorithm uptake, as can the need to secure ethical approval and data sharing agreements. Finally, data ownership should be clear[66] and privacy concerns should be addressed.[67]

Several professional standards for medical AI, such as Decide-AI,[68] STARD-AI,[69] PRIME,[70] TRIPOD-AI, and PROBOST-AI,[71] are shaping the landscape to address limitations in current AI tools and potentially streamline clinical implementation by ensuring quality and reliability in the AI tools being developed and assessed. Decide-AI enhances transparency and interpretability, which fosters trust among clinicians; STARD-AI provides standardised reporting guidelines for diagnostic AI studies, which aids clinicians in assessing reliability; PRIME focuses on robust validation and evaluation of pretrained AI models, particularly in cardiac imaging; and TRIPOD-AI and PROBOST-AI offer further guidance on enhancing the transparency, quality, and accuracy of AI-powered prediction models and risk of bias evaluation. These standards, along with regulatory and ethical considerations[72–74] by the European Commission and US FDA, collectively promote AI's safe and effective integration in health care.

The EU and US FDA have made strides in establishing regulatory frameworks and action plans for AI-based software as a medical device, which focus on evaluation methods, image noise, failure modes, trustworthiness, and generalisability.[75] However, current regulatory standards for AI often struggle to keep pace with rapid technological advancements. Additionally, the absence of standardised evaluation methods for AI technologies complicates regulatory assessments. Furthermore, substantial differences persist in regulatory mechanisms between countries. There is a need for regulatory agencies to collaborate internationally, incorporating bioethical considerations while ensuring comprehensive training for health-care professionals, and active involvement of patients and providers in the development, implementation, and assessment of regulatory frameworks for AI-based and machine learning-based software as a medical device.

Integration and implementation challenges

The scarcity of data on cost-effectiveness and return on investment of AI models, the absence of clear reimbursement models for AI-enabled services, and outdated IT infrastructure hinder the seamless integration of AI applications with existing systems such as picture archiving and communications and electronic health records.[76] Legacy systems are often missing the necessary interfaces and standards for seamless integration, communication, and data exchange, complicating deployment and scalability. Addressing these challenges requires collaborative efforts from policy makers, payers, and technology providers to establish clear reimbursement pathways, invest in infrastructure modernisation, and develop interoperability standards, ultimately unlocking the transformative potential of AI in health-care delivery.[77]

Conclusions

The discussions presented in this Series paper on current AI architectures and the barriers to application for cardiac imaging reinforce that at present, AI is not a panacea to achieving high-value cardiac imaging, given the evolving knowledge of cardiac diseases and unresolved complexities of real-world settings. To make any impact beyond automation, AI tools need to influence and improve clinical decision making and patient outcomes, which will necessitate assessing their efficacy and cost-effectiveness through multicentric registries and pragmatic trials, focusing on multiple measurements in a clinical—not just isolated—context. One strength of AI techniques is that they provide multiparametric integration of complex imaging data for disease patterning in an individual patient. Yet, their potential for precision phenotyping in cardiac imaging remains underused. The shift towards multiparametric phenotypic assessment and integration into physicians' decision-making processes represents a departure from the current landscape of AI tools in cardiac imaging. Overcoming these drawbacks will also require a framework shift from clinical cardiology's traditional norms, which are still dominated by measurements such as ejection fraction towards a framework where cardiac imaging tests open pathways to accurate disease classifications and individualised predictions. We must continue research and development, focusing on overcoming the barriers and carefully integrating AI into existing workflows rather than seeing it as a panacea. While we await new clinical frameworks, AI can be used to reduce the number of low-calibre, less risky, repetitive tasks for physicians (eg, measurements and report

Search strategy and selection criteria

In this Series paper, references for articles written in English were identified through searches of PubMed from Jan 1, 2018 to Dec 31, 2023, with the search terms including "deep learning" or "artificial intelligence" in conjunction with terms related to cardiac imaging techniques: "echocardiography," "cardiac ultrasound," "cardiac computed tomography," "cardiac magnetic resonance," or "nuclear cardiology;" and terms related to clinical trials: "randomized clinical trial," "multicenter prospective," or "prospective multicenter." Articles were also identified through searches of the authors' personal archives.

generations), acting as a form of intellect augmentation, enabling more time to be spent by cardiac imagers and the multidisciplinary teams working with them in conducting crucial inquiries and using creativity in clinical problem-solving. This strategy holds the potential to usher in an era of augmented intelligence, and, in turn, will lay the foundation for efficient and empathetic health-care delivery.

Contributors

PPS conceptualised and developed the paper. All authors collaborated and contributed towards writing the paper. PPS completed the final drafting and assimilation of the paper.

Declaration of interests

DD has received software royalties from and holds a patent with Cedars–Sinai Medical Center. RHD owns shares in Myocardium AI. PPS has served on the Advisory Board of RCE Technologies and HeartSciences and holds stock options; is an Associate Editor for the American College of Cardiology and is a guest editor for the American Society of Echocardiography; received grants or contracts from RCE Technologies, HeartSciences, Butterfly, and MindMics; and hold patents with Mayo Clinic (US8328724B2), HeartSciences (US11445918B2), and Rutgers Health (62/864,771; US202163152686P; WO2022182603A1; US202163211829P; WO2022266288A1; and US202163212228P). NY declares grants or contracts from MindMics, RCE Technologies, HeartSciences, and Abiomed; receives consulting fees from Turnkey Learning and Turnkey Insights; receives payment or honoraria and support for attending meetings or travel from West Virginia University (WVU) and National Science Foundation; is an advisory board member and chair of the student experience committee for Turnkey Learning and Turnkey Insights; is an advisor or board member for Research Spark Hub & Magnetic 3D; is an adjunct professor or faculty member at Carnegie Mellon University; is an editorial board member of American Society of Echocardiography; is a special government employee of the Center for Devices and Radiological Health at the US Food and Drug Association; and holds patents with Rutgers (US202163152686P; WO2022182603A1; US202163211829P; WO2022266288A1; US202163212228P; and WO2022266291A1) and with WVU (invention numbers 2021-20 and 2021-047). ND declares no competing interests.

Acknowledgments

PPS has received funding support from the National Science Foundation (NST) Award (2125872) and the NST Research Traineeship Harnessing the Data Revolution programme: Bridges in Digital Health. DD has received grant funding from the National Institutes of Health and the National Heart, Lung, and Blood Institute (1R01HL148787–01A1 and 1R01HL151266). ND has received partial support from the LABEX PRIMES of Université de Lyon (ANR–11-LABX–0063), the French Agence Nationale de la Recherche (MIC–MAC project and ANR–19-CE45–0005), and the Institut Universitaire de France.

References

1 Dey D, Arnaout R, Antani S, et al. Proceedings of the NHLBI workshop on artificial intelligence in cardiovascular imaging: translation to patient care. *JACC Cardiovasc Imaging* 2023; **16:** 1209–23.

2 Crigger E, Reinbold K, Hanson C, Kao A, Blake K, Irons M. Trustworthy augmented intelligence in health care. *J Med Syst* 2022; **46:** 12.

3 Zhou D-X. Universality of deep convolutional neural networks. *Appl Comput Harmon Anal* 2020; **48:** 787–94.

4 Chen C, Qin C, Qiu H, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med* 2020; **7:** 25.

5 Kissas G, Yang Y, Hwuang E, Witschey WR, Detre JA, Perdikaris P. Machine learning in cardiovascular flows modeling: predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks. *Comput Methods Appl Mech Eng* 2020; **358:** 112623.

6 Pateras J, Rana P, Ghosh P. A taxonomic survey of physics-informed machine learning. *Appl Sci* 2023; **13:** 6892.

7 Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys* 2021; **3:** 422–40.

8 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE. 2016. https://ieeexplore.ieee.org/document/7780459 (accessed March 5, 2024).

9 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.

10 Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006; **313:** 504–07.

11 Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT Press, 2016.

12 van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal* 2022; **79:** 102470.

13 Li J, Chen J, Tang Y, Wang C, Landman BA, Zhou SK. Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives. *Med Image Anal* 2023; **85:** 102762.

14 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *NIPS 17* 2017; **2017:** 6000–10.

15 Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature* 2023; **614:** 214–16.

16 Skalidis I, Cagnina A, Luangphiphat W, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? *Eur Heart J Digit Health* 2023; **4:** 279–81.

17 US Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. May 13, 2024. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (accessed May 13, 2024).

18 Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020; **580:** 252–56.

19 He B, Kwan AC, Cho JH, et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature* 2023; **616:** 520–24.

20 Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol* 2021; **6:** 624–32.

21 Asch FM, Mor-Avi V, Rubenson D, et al. Deep learning-based automated echocardiographic quantification of left ventricular ejection fraction: a point-of-care solution. *Circ Cardiovasc Imaging* 2021; **14:** e012293.

22 Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat Med* 2021; **27:** 882–91.

23 Lancaster MC, Salem Omar AM, Narula S, Kulkarni H, Narula J, Sengupta PP. Phenotypic clustering of left ventricular diastolic function parameters: patterns and prognostic relevance. *JACC Cardiovasc Imaging* 2019; **12:** 1149–61.

24 Sengupta PP, Shrestha S, Kagiyama N, et al. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *JACC Cardiovasc Imaging* 2021; **14:** 1707–20.

25 Pandey A, Kagiyama N, Yanamala N, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. *JACC Cardiovasc Imaging* 2021; **14:** 1887–900.

26 Yacoub B, Varga-Szemes A, Schoepf UJ, et al. Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. *AJR Am J Roentgenol* 2022; **219:** 743–51.

27 Eisenberg E, McElhinney PA, Commandeur F, et al. Deep learning-based quantification of epicardial adipose tissue volume and attenuation predicts major adverse cardiovascular events in asymptomatic subjects. *Circ Cardiovasc Imaging* 2020; **13:** e009829.

28 Choi AD, Marques H, Kumar V, et al. CT evaluation by artificial intelligence for atherosclerosis, stenosis and vascular morphology (CLARIFY): a multi-center, international study. *J Cardiovasc Comput Tomogr* 2021; **15:** 470–76.

29 Lin A, Manral N, McElhinney P, et al. Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *Lancet Digit Health* 2022; **4:** e256–65.

30 Coenen A, Kim YH, Kruk M, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE Consortium. *Circ Cardiovasc Imaging* 2018; **11:** e007217.

31 Yang J, Shan D, Wang X, et al. On-site computed tomography-derived fractional flow reserve to guide management of patients with stable coronary artery disease: the TARGET randomized trial. *Circulation* 2023; **147:** 1369–81.

32 De Backer O, Iriart X, Kefer J, et al. Impact of computational modeling on transcatheter left atrial appendage closure efficiency and outcomes. *JACC Cardiovasc Interv* 2023; **16:** 655–66.

33 Hokken TW, Wienemann H, Dargan J, et al. Clinical value of CT-derived simulations of transcatheter-aortic-valve-implantation in challenging anatomies the PRECISE-TAVI trial. *Catheter Cardiovasc Interv* 2023; published online Sept 5. https://doi.org/10.1002/ccd.30816.

34 Tzimas G, Gulsin GS, Everett RJ, et al. Age- and sex-specific nomographic CT quantitative plaque data from a large international cohort. *JACC Cardiovasc Imaging* 2024; **17:** 165–75.

35 Wang S, Patel H, Miller T, et al. AI based CMR assessment of biventricular function. *JACC Cardiovasc Imaging* 2022; **15:** 413–27.

36 Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* 2018; **20:** 65.

37 Knott KD, Seraphim A, Augusto JB, et al. The prognostic significance of quantitative myocardial perfusion. *Circulation* 2020; **141:** 1282–91.

38 Pieszko K, Shanbhag AD, Singh A, et al. Time and event-specific deep learning for personalized risk assessment after cardiac perfusion imaging. *NPJ Digit Med* 2023; **6:** 78.

39 Singh A, Kwiecinski J, Miller RJH, et al. Deep learning for explainable estimation of mortality risk from myocardial positron emission tomography images. *Circ Cardiovasc Imaging* 2022; **15:** e014526.

40 Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing* 2021; **452:** 48–62.

41 Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. *ICML* 2017; **2017:** 3319–28.

42 Jain S, Wallace BC. Attention is not explanation. *arXiv* 2019; published online Feb 26. https://doi.org/10.48550/arXiv.1902.10186 (preprint).

43 Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. Sanity checks for saliency maps. *NIPS* 2018; **2018:** 9525–36.

44 Saporta A, Gui X, Agrawal A, et al. Benchmarking saliency methods for chest X-ray interpretation. *Nat Mach Intell* 2022; **4:** 867–78.

45 Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digit Med* 2022; **5:** 48.

46 Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv Neural Inf Process Syst* 2022; **35:** 507–20.

47 Kocica MJ, Corno AF, Lačković V, Kanjuh VI. The helical ventricular myocardial band of Torrent-Guasp. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu* 2007; **10:** 52–60.

48 Keulenaer GWD, Brutsaert DL. Systolic and diastolic heart failure are overlapping phenotypes within the heart failure spectrum. *Circulation* 2011; **123:** 1996–2005.

49 Lund LH, Vedin O, Savarese G. Is ejection fraction in heart failure a limitation or an opportunity? *Eur J Heart Fail* 2018; **20:** 431–32

50 Nair A, Tuan LQ, Jones-Lewis N, Raja DC, Shroff J, Pathak RK. Heart failure with mildly reduced ejection fraction—a phenotype waiting to be explored. *J Cardiovasc Dev Dis* 2024; **11:** 148.

51 Duchateau N, Camara O. Chapter 7—Machine learning and biophysical models: how to benefit each other? In: Chinesta F, Cueto E, Payan Y, Ohayon J, eds. Reduced order models for the biomechanics of living organs. London: Elsevier, 2023: 147–64.

52 Armoundas AA, Narayan SM, Arnett DK, et al. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. *Circulation* 2024; **149:** e1028–50.

53 Kitsios GD, Dahabreh IJ, Trikalinos TA, Schmid CH, Huggins GS, Kent DM. Heterogeneity of the phenotypic definition of coronary artery disease and its impact on genetic association studies. *Circ Cardiovasc Genet* 2011; **4:** 58–67.

54 Petersen SE, Matthews PM, Bamberg F, et al. Imaging in population science: cardiovascular magnetic resonance in 100 000 participants of UK Biobank—rationale, challenges and approaches. *J Cardiovasc Magn Res* 2013; **15:** 46.

55 Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. *Patterns* 2021; **2:** 100347.

56 Moon JC, Lorenz CH, Francis JM, Smith GC, Pennell DJ. Breath-hold FLASH and FISP cardiovascular MR imaging: left ventricular volume differences and reproducibility. *Radiol* 2002; **223:** 789–97.

57 Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. *Nat Biomed Eng* 2022; **6:** 1330–45.

58 Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021; **27:** 582–84.

59 Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion* 2021; **76:** 243–97.

60 Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 2020; **21:** 345–52.

61 Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *Int J Hum-Comput Stud* 2021; 146: 102551.

62 Quinn TP, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. *J Am Med Inform Assoc* 2021; **28:** 890–94.

63 Rundo L, Pirrone R, Vitabile S, Sala E, Gambino O. Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine. *J Biomed Inform* 2020; **108:** 103479.

64 Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019; **322:** 1765–66.

65 Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019; **1:** 206–15.

66 Kostkova P, Brewer H, de Lusignan S, et al. Who owns the data? Open data for healthcare. *Front Public Health* 2016; **4:** 7.

67 Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging* 2018; **9:** 745–53.

68 Vasey B, Novak A, Ather S, Ibrahim M, McCulloch P. DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology. *Clin Radiol* 2023; **78:** 130–36.

69 Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open* 2021; **11:** e047709.

70 Sengupta PP, Shrestha S, Berthon B, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare Innovation Council. *JACC Cardiovasc Imaging* 2020; **13:** 2017–35.

71 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.

72 European Commission. Proposal for a regulation laying down harmonised rules on artificial intelligence. April 21, 2021. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (accessed March 11, 2024).

73 European Commission. Ethics guidelines for trustworthy AI. April 8, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed March 11, 2024).

74 US Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). June 13, 2024. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed June 13, 2024).

75 US Food and Drug Administration. FDA releases artificial intelligence/machine learning action plan. Jan 12, 2021. https://www fda gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan (accessed March 11, 2024).

76 Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR and RSNA. *Can Assoc Radiol J* 2024; **75:** 226–44.

77 Seoane F, Traver V, Hazelzet J. Value-driven digital transformation in health and medical care. In: Fernandez-Llatas C, ed. Interactive process mining in healthcare. Cham: Elsevier, 2021: 13–26.