Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints

Ohad Oren, Bernard J Gersh, Deepak L Bhatt

Artificial intelligence (AI) is a disruptive technology that involves the use of computerised algorithms to dissect complicated data. Among the most promising clinical applications of AI is diagnostic imaging, and mounting attention is being directed at establishing and fine-tuning its performance to facilitate detection and quantification of a wide array of clinical conditions. Investigations leveraging computer-aided diagnostics have shown excellent accuracy, sensitivity, and specificity for the detection of small radiographic abnormalities, with the potential to improve public health. However, outcome assessment in AI imaging studies is commonly defined by lesion detection while ignoring the type and biological aggressiveness of a lesion, which might create a skewed representation of AI's performance. Moreover, the use of non-patient-focused radiographic and pathological endpoints might enhance the estimated sensitivity at the expense of increasing false positives and possible overdiagnosis as a result of identifying minor changes that might reflect subclinical or indolent disease. We argue for refinement of AI imaging studies via consistent selection of clinically meaningful endpoints such as survival, symptoms, and need for treatment.

Lancet Digital Health 2020; 2: e486–488

Division of Hematology and Oncology, Mayo Clinic, Rochester, MN, USA (O Oren MD); Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA (Prof B J Gersh DPhil); Brigham and Women's Hospital Heart & Vascular Center and Harvard Medical School, Boston, MA, USA (Prof D L Bhatt MD)

Correspondence to: Prof Deepak L Bhatt, Brigham and Women's Hospital Heart & Vascular Center and Harvard Medical School, Boston, MA 02115, USA
dlbhattmd@post.harvard.edu

The use of artificial intelligence (AI) in diagnostic medical imaging is undergoing extensive evaluation. AI has shown impressive accuracy and sensitivity in the identification of imaging abnormalities and promises to enhance tissue-based detection and characterisation.[1] However, with improved sensitivity emerges an important drawback, namely, the detection of subtle changes of indeterminate significance.[2] For example, an analysis of screening mammograms showed that artificial neural networks are no more accurate than radiologists in detecting cancer—but have consistently higher sensitivity for pathological findings, in particular for subtle lesions.[3] In the beginning of an AI-assisted diagnostic imaging revolution, the medical community has to anticipate the potential unknowns of this technology to ensure effective and safe incorporation into clinical practice. Meticulous assessment of AI's potential perils, in the context of its unique capabilities, is integral to establishing its role in clinical medicine, and navigating between enhanced detection and overdiagnosis will be no easy task. Fundamental to this assessment are consistent use of out-of-sample external validation and well defined cohorts to augment the quality and interpretability of AI studies.[4]

At present, many AI imaging studies estimate diagnostic accuracy by calculating sensitivity and specificity, while others assess clinically important outcomes.[4,5] However, as AI often detects minor image alterations, more relevant outcome variables include new diagnosis of advanced disease, disease requiring treatment, or conditions likely to affect long-term survival. The occurrence of clinically meaningful events—symptoms, need for disease-modifying therapy, and mortality—strongly affect quality of life and should be the focus of AI-based investigations. Even though numerous studies show that AI has higher specificity and lower recall rates than standard reading, such investigations do not typically consider the type and biological aggressiveness of a lesion when estimating accuracy and sensitivity.[6,7] Non-patient-centric endpoint selection might increase sensitivity at the expense of increasing false positives and possibly overdiagnosis as a result of identifying minor changes that could reflect subclinical or indolent disease.

A great challenge is that, unlike discrete findings derived from sophisticated conventional radiographic studies, AI might identify imaging pattern changes that are not easily amenable to human identification.[8,9] For example, analysis of brain MRI using machine learning has the potential to identify tissue changes reflective of early ischaemic stroke within a narrow time window from symptom onset with greater sensitivity than a human reader.[9] Despite the promise of early diagnosis with machine learning, the relationship between very subtle parenchymal brain alterations detected by AI, either in the natural history of small evolving infarcts or non-ischaemic processes, and gross neurological outcomes is unknown. Dedicated studies are needed to ascertain whether AI-defined cerebral changes suggestive of early ischaemia correlate with a different profile of neurologic disability or benefit from thrombolysis. Further, difficult circumstances might ensue in which a recommendation for treatment might be given in the absence of a well defined abnormality detected by routine imaging.[9] At the patient level, such discordance might cause confusion and potentially mistrust and will necessitate public education regarding the new concept of deep learning in imaging analysis. It might also introduce medical liability issues (such as failure to diagnose or potentially unneccessary surgery) that could materialise if AI becomes the standard of care.[10] The public and especially physicians should also be reassured that AI is unlikely to replace radiologists, but a radiologist who uses AI might be more productive than a radiologist who does not.[11,12]

Although these aspects have to be addressed, AI could offer unique opportunities for learning about fine imaging changes reflective of poorly understood disease processes. For example, autoimmune myocarditis is an emerging and potentially fatal complication of immunotherapy.[13] As awareness of this immune-related toxicity increases, cardiac imaging could be done at an earlier timepoint in its natural history, which could potentially lead to an earlier administration of therapy and lower morbidity and mortality. At the same time, a low rule-out threshold will shift the disease phenotype to milder forms of myocarditis. The clinical consequences of low-grade myocardial inflammation in patients who receive immunotherapy will be important to understand. AI could help delineate myocardial tissue changes that reflect inflammation and identify imaging patterns that are highly predictive of treatment response.[14] Harnessing the potential of AI would entail identifying MRI patterns associated with hard clinical outcomes, such as severe arrhythmias, haemodynamic instability, and event-specific mortality, rather than a non-specific but widely encompassing diagnosis of myocarditis. Identification of subtle structural and functional cardiac abnormalities with important clinical correlation could also be accomplished by AI techniques, such as convolutional neural networks, when applied to echocardiography, the most common form of cardiovascular imaging.[15]

Another example comes from the management of patients with aortic stenosis. There is currently no evidence to suggest that patients with non-severe aortic stenosis benefit from valve replacement compared with medical therapy. AI applications of echocardiography, computed tomography, or MRI could provide granular assessment of annular conformation, leaflet mobility, and outflow tract to identify patients with less severe stenosis in whom surgical or percutaneous intervention might be more advantageous than medical management.[16] It is equally and perhaps even more important to identify changes in left ventricular function and fibrosis or remodelling that could play a crucial role in prompting earlier intervention. The enhanced reading performance of AI could be exploited to improve patient selection for intervention by identifying mild structural or dynamic changes that correlate with worse outcomes. A pivotal point is to have accurate AI classification of aortic stenosis severity based on clinically validated input, allowing generation of new observations in a manner congruent with disease phenotype, so that patients with severe disease are correctly captured and those with mild disease are not erroneously reclassified into a high-risk group.

Another high-yield niche for AI imaging is cancer detection and characterisation. High-power quantitative analysis of fine structural image alterations could be used to predict the odds of malignancy and anticipated tumour kinetics and help tailor management plans. A case in point is prostate cancer, which, despite being the most prevalent neoplasm in men, lacks an effective screening strategy. In the past 5 years, multiparametric MRI was shown to increase the detection of clinically relevant prostate cancer, but interobserver variability remains a major obstacle.[17] Deep learning algorithms could enhance the assessment of MRI features such as texture, volume, and shape, and potentially augment the physicians' ability to diagnose advanced prostate cancer, while decreasing biopsies in low-probability cases.[18] The approach for adrenal incidentalomas could also benefit from AI-based imaging analysis. Adrenal nodules are the most frequently encountered incidental radiographic finding and can reflect malignant (ie, pheochromocytoma) or benign (ie, adenoma) conditions with overlapping imaging characteristics.[19] Quantitative texture analysis through high-throughput extraction might differentiate radiographic adrenal lesions into discrete clinical subsets, reducing costly and invasive testing.[20] Replication of AI-guided algorithms in other cancer types would be conducive to generating an unbiased, low-variance machinery for patient-focused imaging interpretation.

The rise and dissemination of AI in clinical medicine will refine our diagnostic accuracy and rule-out capabilities. However, unless AI algorithms are trained to distinguish between benign abnormalities and clinically meaningful lesions, better imaging sensitivity might come at the cost of increased false positives, as well as perplexing scenarios whereby AI findings are not associated with outcomes. To facilitate the study of AI in medical image interpretation, it is paramount to assess the effects on clinically meaningful endpoints to improve applicability and allow effective deployment into clinical practice.

Contributors

OO was responsible for the conceptualisation of this Viewpoint. OO, BJG, and DLB were responsible for the methodology and supervision as well as the writing, reviewing, and editing of the manuscript.

Declaration of interests

BJG discloses the following relationships: Clinical Research Organization for Trials involving Edwards Percutaneous Valve Devices: Baim Institute; Data Safety Monitoring Board—REPRISE study: Boston Scientific Corporation; Data Safety Monitoring Board—RELIEVE-HF & SPYRAL trials: Cardiovascular Research Foundation; Data Safety Monitoring Board—Pioneer HCM trail: Duke Clinical Research Institute; Data Safety Monitoring Board: Duke University; Data Safety Monitoring Board—ENVISAGE-TAVI trial: Icahn School of Medicine at Mount Sinai; Steering Committee—REVEAL trial: Medtronic; Data Safety Monitoring Board—PROMINENT trial: Kowa Research Institute; General Consulting: MyOKardia; Executive Committee ORBIT Registries: Janssen Scientific Affairs. DLB discloses the following relationships, outside the submitted work: grants from Amarin; grants from AstraZeneca; grants from Bristol-Myers Squibb; grants from Eisai; grants from Ethicon; grants from Medtronic; grants from Sanofi-Aventis; grants from The Medicines Company; other from FlowCo; grants and other from PLx Pharma; other from Takeda; personal fees from Duke Clinical Research Institute; personal fees from Mayo Clinic; personal fees from Population Health Research Institute; personal fees, non-financial support, and other from American College of Cardiology; personal fees from Belvoir Publications; personal fees from Slack Publications; personal fees from WebMD; personal fees from Elsevier; other from Medscape Cardiology; other from Regado

Biosciences; other from Boston VA Research Institute; personal fees and non-financial support from Society of Cardiovascular Patient Care; non-financial support from American Heart Association; personal fees from HMP Global; grants from Roche; personal fees from Harvard Clinical Research Institute (now Baim Institute for Clinical Research); other from Clinical Cardiology; personal fees from Journal of the American College of Cardiology; other from Veterans Affairs; grants from Pfizer; grants from Forest Laboratories/AstraZeneca; grants from Ischemix; other from St Jude Medical (now Abbott); other from Biotronik; grants and other from Cardax; other from Boston Scientific; grants from Amgen; grants from Lilly; grants from Chiesi; grants from Ironwood; personal fees from Cleveland Clinic; personal fees from Mount Sinai School of Medicine; other from Merck; grants from Abbott; grants from Regeneron; other from Svelte; grants and other from PhaseBio; grants from Idorsia; grants from Synaptic; personal fees from TobeSoft; grants, personal fees, and other from Boehringer Ingelheim; personal fees from Bayer; other from Novo Nordisk; grants from Fractyl; personal fees from Medtelligence/ReachMD; personal fees from CSL Behring; other from Cereno Scientific; grants from Afimmune; personal fees from Ferring Pharmaceuticals; other from Cardiovascular Systems; grants from Lexicon; personal fees from MJH Life Sciences; personal fees from Level Ex; personal fees from Contego Medical; personal fees from CellProthera; and personal fees from K2P. OO has no competing interests.

References

1 Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digital Health* 2020; **2:** e138–48.

2 van den Heuvel TL, van der Eerden AW, Manniesing R, et al. Automated detection of cerebral microbleeds in patients with traumatic brain injury. *Neuroimage Clin* 2016; **12:** 241–51.

3 Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Invest Radiol* 2017; **52:** 434–40.

4 Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health* 2019; **1:** e271–97.

5 Bello GA, Dawes TJW, Duan J, et al. Deep learning cardiac motion analysis for human survival prediction. *Nat Mach Intell* 2019; **1:** 95–104.

6 Park VY, Han K, Seong YK, et al. Diagnosis of thyroid nodules: performance of a deep learning convolutional neural network model vs. radiologists. *Sci Rep* 2019; **9:** 17843.

7 Jiang H, Ma H, Qian W, Gao M, Li Y. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. *IEEE J Biomed Health Inform* 2018; **22:** 1227–37.

8 Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med* 2018; **1:** 9.

9 Lee H, Lee EJ, Ham S, et al. Machine learning approach to identify stroke within 4·5 hours. *Stroke* 2020; **51:** 860–66.

10 Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019; **18:** 1765–66.

11 Oren O, Kebebew E, Ioannidis JPA. Curbing unnecessary and wasted diagnostic imaging. *JAMA* 2019; **321:** 245–46.

12 Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* 2018; **1:** 35.

13 Ball S, Ghosh RK, Wongsaengsak S, et al. Cardiovascular toxicities of immune checkpoint inhibitors: JACC review topic of the week. *J Am Coll Cardiol* 2019; **74:** 1714–27.

14 Baessler B, Luecke C, Lurz J, et al. Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology* 2018; **289:** 357–65.

15 Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med* 2020; **3:** 10.

16 Kwon JM, Lee SY, Jeon KH, et al. Deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc* 2020; **9:** e014717.

17 Stabile A, Giganti F, Rosenkrantz AB, et al. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. *Nat Rev Urol* 2020; **17:** 41–61.

18 Yoo S, Gujrathi I, Haider MA, Khalvati F. Prostate cancer detection using deep convolutional neural networks. *Sci Rep* 2019; **9:** 19518.

19 Oren O, Blankstein R, Bhatt DL. Incidental imaging findings in clinical trials. *JAMA* 2020; **7:** 603–04.

20 Yi X, Guan X, Chen C, et al. Adrenal incidentaloma: machine learning-based quantitative texture analysis of unenhanced CT can effectively differentiate sPHEO from lipid-poor adrenal adenoma. *J Cancer* 2018; **19:** 3577–82.