










AI and machine learning in medical imaging: key points from development to translation

Ravi K. Samala , PhD^{*1}, Karen Drukker, PhD², Amita Shukla-Dave , PhD^{3,4}, Heang-Ping Chan , PhD⁵, Berkman Sahiner, PhD¹, Nicholas Petrick, PhD¹, Hayit Greenspan, PhD⁶, Usman Mahmood , PhD⁴, Ronald M. Summers , MD, PhD⁷, Georgia Tourassi, PhD⁸, Thomas M. Deserno , PhD⁹, Daniele Regge, MD^{10,11}, Janne J. Näppi, PhD¹², Hiroyuki Yoshida , PhD¹², Zhimin Huo, PhD¹³, Quan Chen, PhD¹⁴, Daniel Vergara, MS¹⁵, Kenny H. Cha , PhD¹, Richard Mazurchuk, PhD¹⁶, Kevin T. Grizzard, PhD¹⁷, Henkjan Huisman , PhD¹⁸, Lia Morra, PhD¹⁹, Kenji Suzuki, PhD²⁰, Samuel G. Armato, III, PhD², Lubomir Hadjiiski, PhD⁵

¹Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, 20993, United States

²Department of Radiology, University of Chicago, Chicago, IL, 60637, United States

³Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065, United States

⁴Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065, United States

⁵Department of Radiology, University of Michigan, Ann Arbor, MI, 48109, United States

⁶Biomedical Engineering and Imaging Institute, Department of Radiology, Icahn School of Medicine at Mt Sinai, New York, NY, 10029, United States

⁷Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD, 20892, United States

⁸Computing and Computational Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, United States

⁹Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Braunschweig, Niedersachsen, 38106, Germany

¹⁰Radiology Unit, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, 10060, Italy

¹¹Department of Translational Research and of New Surgical and Medical Technologies of the University of Pisa, Pisa, 56126, Italy

¹²3D Imaging Research, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114, United States

¹³Tencent America, Palo Alto, CA, 94306, United States

¹⁴Department of Radiation Oncology, Mayo Clinic Arizona, Phoenix, AZ, 85054, United States

¹⁵Department of Radiology, University of Washington, Seattle, WA, 98195, United States

¹⁶Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, United States

¹⁷Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, CT, 06510, United States

¹⁸Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Gelderland, 6525 GA, Netherlands

¹⁹Department of Control and Computer Engineering, Politecnico di Torino, Torino, Piemonte, 10129, Italy

²⁰Institute of Innovative Research, Tokyo Institute of Technology, Midori-ku, Yokohama, Kanagawa, 226-8503, Japan

*Corresponding author: Ravi K. Samala, PhD, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, 20993, United States (Ravi.Samala@fda.hhs.gov)

Abstract

Innovation in medical imaging artificial intelligence (AI)/machine learning (ML) demands extensive data collection, algorithmic advancements, and rigorous performance assessments encompassing aspects such as generalizability, uncertainty, bias, fairness, trustworthiness, and interpretability. Achieving widespread integration of AI/ML algorithms into diverse clinical tasks will demand a steadfast commitment to overcoming issues in model design, development, and performance assessment. The complexities of AI/ML clinical translation present substantial challenges, requiring engagement with relevant stakeholders, assessment of cost-effectiveness for user and patient benefit, timely dissemination of information relevant to robust functioning throughout the AI/ML lifecycle, consideration of regulatory compliance, and feedback loops for real-world performance evidence. This commentary addresses several hurdles for the development and adoption of AI/ML technologies in medical imaging. Comprehensive attention to these underlying and often subtle factors is critical not only for tackling the challenges but also for exploring novel opportunities for the advancement of AI in radiology.

Keywords: medical imaging; artificial intelligence/machine learning; AI/ML considerations; performance evaluation; AI/ML lifecycle.

Introduction

The rise of AI for medical imaging has sparked concerns about the potential obsolescence of radiology as a profession.

This commentary paper seeks to provide a more balanced perspective on the near-future role of AI in clinical radiology and explores the challenges in AI development and

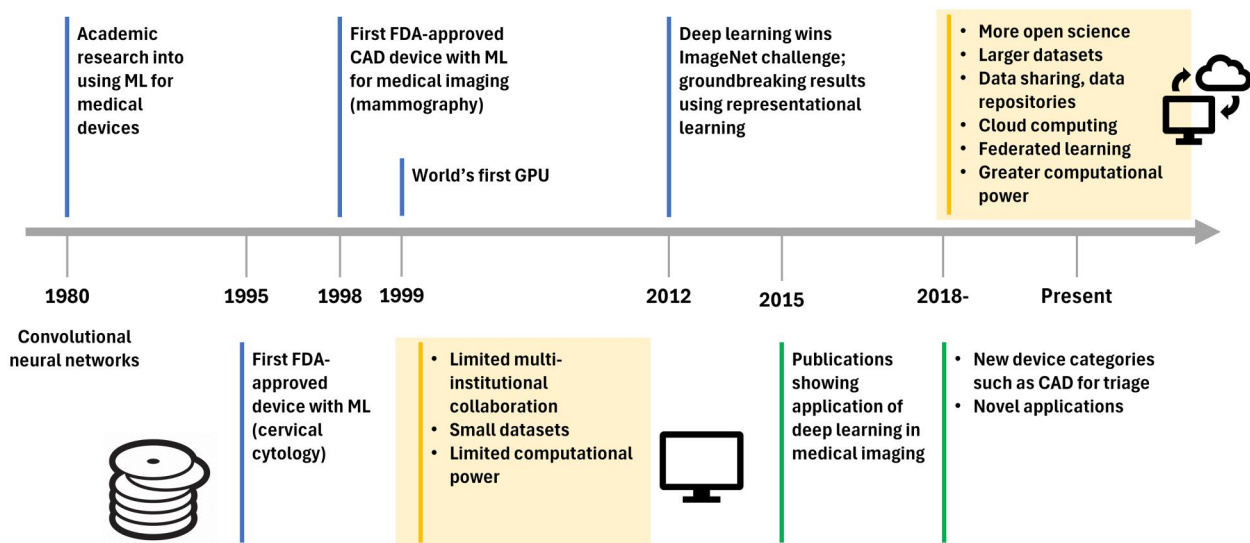


Figure 1. Journey of AI/ML in medical imaging.

implementation despite proliferation of AI systems. Prior to the early 2000s, limited data sets, insufficient computational power, and siloed research groups hindered progress (Figure 1). Nonetheless, conventional AI/ML algorithms were successfully implemented in the radiology and healthcare domains, with notable applications for breast cancer in mammography, lung nodules in chest CT, and quantitative coronary angiography. Breakthrough work in 2012 established the potential of novel deep learning (DL) architectures; coupled with the availability of modern graphics processing units and large datasets in the computer vision field, these advances accelerated the progress of AI/ML development, which propagated into medical imaging as early as 2015.¹

The first U.S. FDA-approved medical device, which used an ML algorithm for the interpretation of cervical cytology slides in a semi-automated manner, was approved in 1995 (U.S. Food and Drug Administration. PAPNET testing system: summary of safety and effectiveness. https://www.accessdata.fda.gov/cdrh_docs/pdf/p940029.pdf 1995; P950009), and the first computer-aided diagnosis device for medical imaging, developed for mass and microcalcification detection in mammography, was approved for clinical use in 1998 (U.S. Food and Drug Administration. M1000 IMAGECHECKER: summary of safety and effectiveness. https://www.accessdata.fda.gov/cdrh_docs/pdf/p970058.pdf. 1998; P970058). Since 2018, there have been many AI/ML-enabled medical devices for different clinical tasks and across different types of data acquisition systems due to open science initiatives, availability of large data sets, algorithmic improvements, and easier accessibility. Consequently, AI/ML applications in radiology accounted for an impressive 87% of devices authorized by the U.S. FDA in 2022 (122 devices) (artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices, accessed in January 2024: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>). The journey from initial AI/ML development to regulatory approval to clinical adoption has been the topic of much ongoing research and many publications (Table 1).

Despite tremendous progress in medical imaging AI/ML, important challenges remain that, if not thoroughly

investigated and addressed, may negatively impact patient care, raise patient privacy concerns, and expose other ethical issues such as amplification of inequities in healthcare. The following sections explore important considerations in (1) design and development, (2) performance assessment, and (3) translation phases of the AI/ML lifecycle (Figure 2), factors that impact the ability of AI/ML systems to operate independently in the near future, whether within radiology or within the broader healthcare ecosystem.

Design and development

In recent years, medical imaging research has experienced a shift towards greater openness by embracing data and code sharing, inter-disciplinary collaboration, scientific grand challenges, collaborative platforms such as GitHub, and software platforms for fast prototyping of AI/ML algorithms. The new NIH data management and sharing policy may help further advance data sharing.¹³ Data quality and quantity for AI/ML training and testing, however, remain among the most important impediments to developing AI/ML for many critical clinical applications. Insufficient quantity of high-quality data also has major consequences in limiting model generalizability, equity, and fairness. Regardless of the characteristics of the development datasets, reliable and trustworthy performance requires that algorithm validation use well-curated, high-quality, previously unseen external data that are representative of the intended use and patient population.

Subgroup and data imbalances during model development can hinder AI/ML model performance and generalizability. Imbalances in the data can be broadly categorized into three groups: (1) demographic differences such as age, sex, race, and ethnicity and patient-specific characteristics such as disease severity and underlying conditions; (2) factors dependent on the data collection, image acquisition, and data conversion processes; and (3) model development practices such as use of pretrained models that may propagate biases. Traditional mitigation approaches based on data sampling, cost-sensitive learning using custom loss functions, and post-hoc calibration may address these issues; however, these approaches need to be investigated further for the modern

Table 1. Literature providing additional information on the considerations listed in Figure 2. A search in Scopus for review articles from 2018-present on AI in medical imaging resulted in 94 general review articles. The selection of the 11 reviews listed here is based on the consensus of authors to cover the main topics of the present opinion paper.

Study reference	Title	Objective	Conclusion
Balki et al ²	Sample-size determination methodologies for machine learning in medical imaging research: A systematic review	To provide a descriptive review of current sample-size determination methodologies in ML applied to MI and to propose recommendations for future work in the field.	There is scarcity of research in training set size determination methodologies applied to ML in MI, a need for work in development and streamlining of pre-hoc and post-hoc sample size approaches, and a need to standardize current reporting practices.
El Naqa et al ³	Artificial intelligence: Reshaping the practice of radiological sciences in the 21 st century	To reflect on lessons learned from AI's past history and summarize the current status of AI in radiological sciences, highlighting its impressive achievements and effect on re-shaping the practice of MI and radiotherapy in the areas of computer-aided detection, diagnosis, prognosis, and decision support.	A summary of the current challenges to overcome for AI to achieve its promise of providing better precision healthcare for each patient while reducing cost burden on their families and the society at large.
England and Cheng ⁴	Artificial intelligence for medical image analysis: A guide for authors and reviewers	To highlight best practices for writing and reviewing articles on AI for medical image analysis.	AI/ML is in the early phases of application to MI, and patient safety demands a commitment to sound methods and avoidance of rhetorical and overly optimistic claims. Adherence to best practices should elevate the quality of articles submitted to and published by clinical journals.
Huff et al ⁵	Interpretation and visualization techniques for deep learning models in medical imaging	To summarize currently available methods for performing image model interpretation and critically evaluate published uses of these methods for MI applications.	Recap of toolkits for model interpretation specific to MI applications, limitations of current model interpretation methods, recommendations for DL practitioners looking to incorporate model interpretation into their task, and a general discussion on the importance of model interpretation in MI contexts.
Pesapane et al ⁶	Artificial intelligence as a medical device in radiology: Ethical and regulatory issues in Europe and the United States	To analyze the legal framework regulating medical devices and data protection in Europe and in the United States, assessing ongoing developments.	The processes of medical device decision-making are largely unpredictable, therefore holding the creators accountable clearly raises concerns. There is much that can be done to regulate (AI/ML)-enabled applications and, if done properly, the potential of AI/ML-based technology, in radiology as well as in other fields, will be invaluable.
Reyes et al ⁷	On the interpretability of artificial intelligence in radiology: Challenges and opportunities	To provide insights into the current state of the art for interpretability methods for radiology AI/ML.	Radiologists' opinions on the topic and of trends and challenges that need to be addressed to effectively streamline interpretability methods in clinical practice were presented.
Sahiner et al ⁸	Deep learning in medical imaging and radiation therapy	To (a) summarize what has been achieved to date in AI/ML for MI; (b) identify challenges and strategies that researchers have taken to address these challenges; and (c) identify promising avenues for future applications and technical innovations.	Lessons learned, remaining challenges, and future directions were presented.
van der Velden et al ⁹	Explainable artificial intelligence (XAI) in deep learning-based medical image analysis	To provide an overview of XAI used in DL-based medical image analysis.	There are many future opportunities for XAI in medical image analysis.

(continued)

Table 1. (continued)

Study reference	Title	Objective	Conclusion
Willeminck et al ¹⁰	Preparing medical imaging data for machine learning	To describe fundamental steps for preparing MI data in AI/ML algorithm development, explain current limitations in data curation, and explore new approaches to address the problem of data availability.	The potential applications are vast and include the entirety of the MI life cycle from image creation to diagnosis to outcome prediction, but existing limitations need to be overcome.
Hadjiiski et al ¹¹	AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging	To bring attention to the proper training and validation of ML algorithms that may improve their generalizability and reliability and accelerate the adoption of CAD-AI systems for clinical decision support.	Rigor and reproducibility of CAD-AI systems will provide the foundation for the success of such systems when translated into clinical practice. Following the best practices during system development and validation should increase the chance of clinical translation for the developed system.
Tang et al ¹²	Canadian Association of Radiologists white paper on artificial intelligence in radiology	To provide recommendations derived from deliberations among members of the Association's AI working group.	Key terminology, educational needs, research and development, partnerships, potential clinical applications, implementation, structure and governance, role of radiologists, and potential impact of AI on radiology in Canada were presented.

Abbreviations: MI = medical imaging, AI = artificial intelligence, ML = machine learning, DL = deep learning, XAI = explainable artificial intelligence.

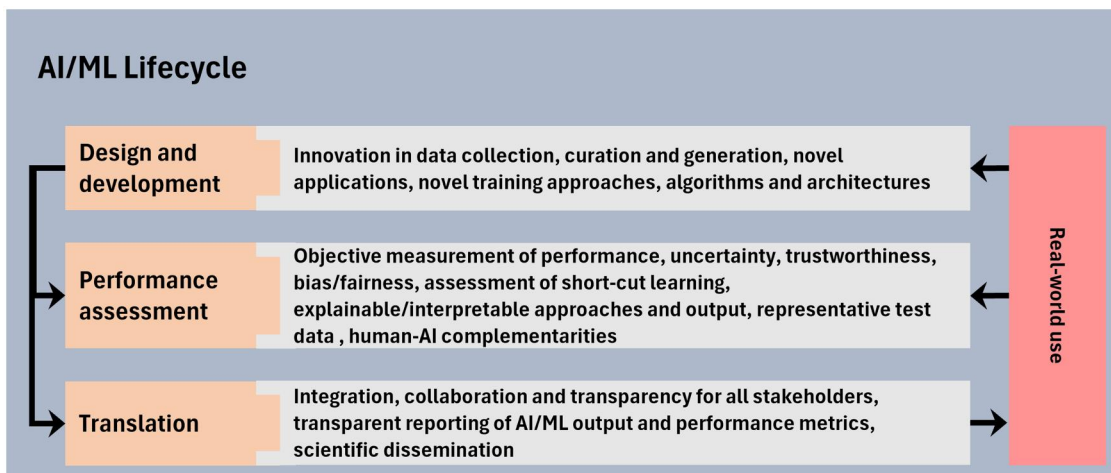


Figure 2. Important considerations in the advancement and translation of AI/ML-enabled medical imaging applications.

data-driven, deep-learning architectures. In addition, with the expansion of AI into the clinical workflow, appropriate care must be taken to understand the consequences of false negatives and false positives based on the intended use of the AI/ML models.

Several approaches have been proposed to improve model generalizability by addressing data imbalances or dataset shift, including the use of privacy-preserving federated learning, and continual learning models; while these approaches are promising, many of the underlying issues remain unsolved. Privacy-preserving federated learning has the potential to train AI/ML models with more data representing greater diversity, which could help improve model generalizability and overall performance. However, differential

privacy (a concept that safeguards individual privacy by ensuring that the inclusion or exclusion of a single person's data is not traceable) is difficult to guarantee.^{14,15} Continual learning algorithms allow model weights to evolve over time, thus providing an approach that adapts to drifts in the input data so that the AI/ML tool can maintain or even improve performance over time. Self-adapting and evolving algorithms such as those with continual learning capabilities could result in a steady improvement in model performance, but a stringent quality assurance program should be implemented to continuously monitor and validate their performance in the clinical environment.

Synthetic datasets generated using generative AI approaches or *in silico* models have the potential to reduce

the imbalances in development datasets. However, there is still a need to innovate generative AI methodologies to reduce the creation of artefacts, generate higher-resolution images, and ensure realistic output. These limitations could reduce AI/ML utility by negatively affecting performance and creating safety issues in downstream tasks. For example, generative AI-assisted image reconstruction could introduce artefacts that hinder human interpretation tasks or affect the robustness of downstream AI/ML-aided detection tasks. Rigorous validation with real-world patient data is needed before AI/ML models developed with or without the assistance of synthetic data can be adopted for clinical use.

An important aspect of overall data quality is the integrity of their annotations (labels) or reference standard. With recent advances in large language models such as ChatGPT, the automatic or semi-automatic generation of reference annotations based on information extracted from electronic health records through natural language processing has become increasingly possible for training AI/ML algorithms. However, the quality of annotations may be suboptimal compared with manual curation by domain experts. The tradeoff between a larger training sample with ‘noisy’ labels and a smaller sample with cleaner labels is a topic of continued investigation.

New AI/ML device categories such as computer-aided triage, acquisition, optimization, and denoising have resulted in novel applications, but an overarching problem with AI/ML algorithms is the lack of understanding of their decision-making processes. The absence of interpretability not only undermines trust among radiologists and clinicians but also diminishes patient trust in AI/ML. Moreover, AI is not neutral: AI/ML decisions, like clinician decisions, are susceptible to inaccuracies, discriminatory outcomes, and embedded or inserted biases. For example, it is known that “shortcut learning”, in which an AI/ML model learns to make decisions based on characteristics other than those intended (eg, output based on the presence of breathing tubes rather than on COVID-19 severity), is not only detrimental to generalizability but can also amplify biases present in the training data. Proper bias identification and mitigation methods should be employed to minimize these problems before they cause generalization, and potentially health inequity, issues.

Performance evaluation

As radiologic AI/ML-enabled applications continue to expand, new evaluation approaches may be required.¹⁶ Additional research is needed to understand human-AI interactions for medical imaging that broadly includes complementarity, transparency, and interpretability. Systematic understanding of the impact of AI on human performance and tuning the AI to complement the clinician through the use of novel frameworks¹⁷ are needed to make the optimum use of AI/ML tools, including seamless integration and improvement in human-perceived subjective factors such as performance, trustworthiness, and usability.

To establish the trustworthiness of AI, it is essential to clarify the relationship between input and AI output. Achieving this necessitates additional research and the application of mechanistic interpretability (reverse engineering fundamental AI/ML model components for human understanding). Uncertainty associated with the AI/ML device outputs should be an essential part of scientific communication and should be reported in a manner intuitive to clinicians. Consensus on

types of uncertainty and uncertainty estimation in AI/ML models and methods remains elusive. Explainability is particularly challenging in open-set AI clinical applications and monitoring where the model encounters data outside the predefined classes used for training. In addition, safe and trustworthy deployment requires detection, characterization, and mitigation of distribution shifts.

Implementing quality checks for scientific manuscripts is another ongoing challenge. Initiatives such as STARD,¹⁸ journal editorial guidelines,¹⁹ and recommended best practices from an American Association of Physicists in Medicine task group¹¹ aim to improve the validity and generalizability of reported performance of AI/ML algorithms. These guidelines promote rigorous study design, appropriate selection of endpoints and metrics, and the use of reliable test data as well as consideration of bias, equity, and fairness.

Conventional strategies for performance evaluation of AI/ML models are insufficient to provide adequate assessment of generalizability, trustworthiness, uncertainty, and explainability. A system that passes all conventional tests in a controlled environment nevertheless may not be ready for clinical use. Innovations in evaluation strategies and further testing during the acceptance phase are essential.

Translation

The application of AI/ML algorithms in clinical care requires rigorous evaluation, validation, and regulatory approval. AI acceptance by healthcare providers, practitioners, and patients is also inherently tied to the important considerations mentioned in the previous sections. Addressing these challenges, including data quality, algorithm robustness, regulatory considerations, clinical workflow integration, patient safety, and continued quality assurance in an efficient manner, is crucial. Collaboration among teams of stakeholders that represent AI/ML developers, regulators, and local AI/ML domain experts along with physicists, clinicians, hospital administrators, payors, and patient advocacy groups is needed to aid in the transition of an AI/ML device from development to clinical practice.²⁰

Transparency is defined as the degree to which appropriate information about a device, including its intended use, development, performance, and, when available, logic, is clearly communicated to stakeholders (U.S. Food and Drug Administration. Virtual Public Workshop—Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices. October 14, 2021. https://downloads.regulations.gov/FDA-2019-N-1185-0138/attachment_1.pdf).

Determining what information is necessary, the mode of communication, and the target user is a practical challenge with implications for improved patient safety and for overcoming automation bias due to human-AI interactions.²¹ Transparency is a persistent concern throughout all three stages of the AI/ML lifecycle, a crucial aspect that cannot be overstated. Its absence hampers the translation of AI/ML technologies into clinical practice and impedes their acceptance as supportive tools for radiologists. Lack of transparency also may prolong the time until AI/ML devices are capable of autonomous deployment in the clinic. During the translation phase, transparency is essential for monitoring AI/ML performance in clinical settings, identifying potential biases in patient subgroup performance, and detecting any deviations from performance observed during development.

The diverse nature of healthcare delivery across organizations and the varied applications of AI/ML poses significant challenges in establishing a comprehensive framework of quality assurance, quality control, and acceptance testing to ensure consistent safety and effectiveness. Additionally, unresolved questions about legal liability persist when AI systems make errors. Such a framework, which is expected to influence every stage from development to translation, must be achieved expediently, especially as the field of medical imaging stands on the brink of the AI revolution. Nonetheless, as we have argued here, this transformation may not occur as rapidly or with as extensive an impact as some AI advocates suggest. Collaborative approaches to quality assurance may be part of a solution, whereby healthcare organizations work together to develop practical QA methods and share data to enable effective risk assessment and mitigation strategies for specific AI tools.

Conclusion

It is anticipated that many AI/ML-enabled devices will be integrated into multiple facets of the clinical workflow for medical imaging in the foreseeable future. To date, these devices have largely served as a second opinion for decision support. However, there is an expectation that patient care may be further improved if certain AI applications can operate autonomously in a trustworthy manner. Nonetheless, the challenges associated with developing robust AI/ML applications for medical imaging, and the work that remains to resolve these challenges, reduce concerns regarding perceived threats to the radiology profession; instead, a clinician-AI partnership or symbiosis is emerging as a more practical and realistic goal. The successful translation of AI/ML into clinical practice in medical imaging demands a critical perspective, enabling the identification and resolution of challenges while uncovering new opportunities for advancement.

Funding

Karen Drukker – supported in part by MIDRC (The Medical Imaging and Data Resource Center) through the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contract 75N92020D00021. Heang-Ping Chan – supported in part by the National Institutes of Health Award Number R01CA214981. Hayit Greenspan was supported by grant UL1TR004420 from the National Center for Advancing Translational Sciences, National Institutes of Health. Ronald M Summers – supported by the Intramural Research Program of the National Institutes of Health Clinical Center. Janne J. Näppi – supported by National Institutes of Health (NIH) Grant Nos. R01CA212382, R01HL164697, and the Interim Support Funding of the Massachusetts General Hospital Executive Committee on Research (ECOR). Hiroyuki Yoshida – supported by National Institutes of Health (NIH) Grant Nos. R01CA212382, R01HL164697, and the Interim Support Funding of the Massachusetts General Hospital Executive Committee on Research (ECOR). Samuel G. Armato III – supported in part by MIDRC (The Medical Imaging and Data Resource Center) through the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contract 75N92020D00021. Lubomir Hadjiiski was supported by the

National Institutes of Health Award Number U01-CA232931.

Conflicts of interest

Ravi K. Samala – nothing to disclose. Karen Drukker – receives royalties from Hologic. Amita Shukla-Dave – nothing to disclose. Heang-Ping Chan – nothing to disclose. Berkman Sahiner – nothing to disclose. Nicholas Petrick – nothing to disclose. Hayit Greenspan – nothing to disclose. Usman Mahmood – nothing to disclose. Ronald M Summers – received royalties for patents or software licenses from iCAD, Philips, ScanMed, Translation Holdings, PingAn and MGB, received research support from PingAn through a Cooperative Research and Development Agreement, not related to this work. Georgia Tourassi – nothing to disclose. Thomas M. Deserno – nothing to disclose. Daniele Regge – nothing to disclose. Janne J. Näppi – has received royalties from Hologic and from MEDIAN Technologies, through the University of Chicago licensing, not related to this work. Hiroyuki Yoshida – has received royalties from licensing fees to Hologic and Medians Technologies through the University of Chicago licensing, not related to this work. Zhimin Huo – nothing to disclose. Quan Chen – has received compensations from Carina Medical LLC, not related to this work, provides consulting services for Reflexion Medical, not related to this work. Daniel Vergara – nothing to disclose. Kenny Cha – nothing to disclose. Richard Mazurchuk – nothing to disclose. Kevin T. Grizzard – nothing to disclose. Henkjan Huisman – has received grant support from Siemens Healthineers and Canon Medical for a scientific research project, not related to this work. Lia Morra – has received funding from HealthTriagesrl, not related to this work. Kenji Suzuki – provides consulting services for Canon Medical, not related to this work. Samuel G. Armato III – has received royalties and licensing fees for computer-aided diagnosis through the University of Chicago, not related to this work.

References

- Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans Med Imaging*. 2016;35(5):1153-1159.
- Balki I, Amirabadi A, Levman J, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Can Assoc Radiol J*. 2019;70(4):344-353. <https://doi.org/10.1016/j.carj.2019.06.002>
- El Naqa I, Haider MA, Giger ML, Ten Haken RK. Artificial intelligence: reshaping the practice of radiological sciences in the 21st century. *Br J Radiol*. 2020;93(1106):e20190855. <https://doi.org/10.1259/bjr.20190855>
- England JR, Cheng PM. Artificial intelligence for medical image analysis: a guide for authors and reviewers. *AJR Am J Roentgenol*. 2019;212(3):513-519. <https://doi.org/10.2214/AJR.18.20490>
- Huff DT, Weisman AJ, Jeraj R. Interpretation and visualization techniques for deep learning models in medical imaging. *Phys Med Biol*. 2021;66(4):04tr01. <https://doi.org/10.1088/1361-6560/ab cd17>
- 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(5):745-753. <https://doi.org/10.1007/s13244-018-0645-y>
- Reyes M, Meier R, Pereira S, et al. On the interpretability of artificial intelligence in radiology: Challenges and opportunities. *Radiol*

- Artif Intell.* 2020;2(3):e190043. <https://doi.org/10.1148/ryai.2020190043>
8. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys.* 2019;46(1):e1-e36. <https://doi.org/10.1002/mp.13264>
 9. 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. <https://doi.org/10.1016/j.media.2022.102470>
 10. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Review. Radiology.* 2020;295(1):4-15. <https://doi.org/10.1148/radiol.2020192224>
 11. Hadjiiski L, Cha K, Chan HP, et al. AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging. *Med Phys.* 2023;50(2):e1-e24.
 12. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J.* 2018;69(2):120-135.
 13. National Institutes of Health. NIH Data Sharing; 2023. Accessed August 21, 2023.
 14. Near JP, Darais D, Lefkovitz N, Howarth G. Guidelines for Evaluating Differential Privacy Guarantees. 2023.
 15. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell.* 2020;2(6):305-311.
 16. Drukker K, Chen W, Gichoya J, et al. Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. *J Med Imaging (Bellingham).* 2023;10(6):061104-061104.
 17. Steyvers M, Tejada H, Kerrigan G, Smyth P. Bayesian modeling of human-AI complementarity. *Proc Natl Acad Sci U S A.* 2022;119(11):e2111547119.
 18. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6(11):e012799.
 19. El Naqa I, Boone JM, Benedict SH, et al. AI in medical physics: guidelines for publication. *Med Phys.* 2021;48(9):4711-4714.
 20. Mahmood U, Shukla-Dave A, Chan H-P, et al. Artificial intelligence in medicine: mitigating risks and maximizing benefits via quality assurance, quality control, and acceptance testing. 2024;1(1):ubae003.
 21. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology.* 2023;307(4):e222176.