

The Future of Artificial Intelligence in Interventional Oncology

Potential applications of AI within IO include patient diagnosis, response prediction, and intraprocedural guidance, dependent upon standardized and responsible implementation.

By Helen Zhang, BS; Shreyas Kulkarni, BS; Zhicheng Jiao, PhD; and Harrison X. Bai, MD, MS

Artificial intelligence (AI) has demonstrated tremendous potential to enhance clinical decision-making, with models capable of identifying complex underlying relationships within data.¹ In the context of image-guided minimally invasive therapies, AI may be able to rapidly interpret multimodality imaging and clinical data to provide personalized clinical support. Although the field of interventional oncology (IO) is rapidly evolving, clinical applications of AI are limited. The adoption of AI within IO has the potential to significantly improve patient diagnosis, treatment, and management.

CURRENT APPLICATIONS OF AI IN IO

Patient Identification and Diagnosis

AI technologies have been prominently developed for patient identification and diagnosis.² An increasingly common approach in cancer detection, radiomics is a method of assessing lesions through image characteristics invisible to the naked eye.³ AI has demonstrated near-physician abilities in detecting and classifying lesions, with improved performance when such tools are used in concert with radiologists.² AI-driven algorithms may also impact how clinicians determine tumor severity. Through advanced radiomics, AI has demonstrated prowess in classifying tumors in hepatocellular carcinomas and assessing malignancy in breast lesions. This AI method of evaluation can eliminate the need for excessive invasive biopsies and reduce health care costs for patients.² As with most AI tasks, these decision-

making models require a significant amount of data, representative of the target population. To be properly integrated into clinical practice and care, institutional protocols for producing representative models must be in place to protect patients and prevent exacerbation of health care inequity.⁴

Patient Selection

AI models have achieved notable performance in patient selection and response prediction compared to existing treatment algorithms and staging systems.⁵⁻⁸ Reliance of traditional systems on a small number of clinical, laboratory, and qualitative imaging features limits the extent of patient characterization possible. However, AI models utilize all of a patient's data to provide more individualized staging. Models have demonstrated strong results in predicting response to procedures such as transarterial chemoembolization and tumor ablation.^{5,9} These algorithms can be employed to identify patients who may be suitable candidates for certain procedures and provide personalized treatment plans.

Intraprocedural Guidance

AI may be able to directly provide guidance during IO procedures, although there is limited literature describing such approaches. A suggested application is image fusion of intraprocedural imaging with diagnostic scans, providing a multimodality approach to real-time tumor localization.^{10,11} Theoretical models may also analyze

relationships between procedural approach and therapeutic effect in real time, providing intraprocedural guidance to optimize patient outcomes. These models may help guide catheter navigation, ablation, probe placement, and other IO techniques. For example, several studies have trained convolutional neural networks to improve needle tip localization.^{12,13} These findings indicate that AI may have many unexplored applications particularly relevant to IO.

Innovations in Large Language Models

There has recently been significant development in the natural language processing space, with chatbots like ChatGPT (OpenAI) and Bard (Google) becoming popular in the public domain.¹⁴ GPT-4 (OpenAI) and Med-PaLM 2 (Google) have performed above 80% on United States medical licensing examinations, prompting discussion of their potential clinical applications.^{15,16} For IO, large language models (LLMs) may be useful for patient education on interventional procedures, producing patient-specific explanations.¹⁷ Additionally, these may serve a role as medical decision-making tools for clinicians; LLMs have produced near-physician results in breast tumor board recommendations.¹⁸ The synthesis ability of LLMs enables them to condense large knowledge bases, potentially providing clinical decision support in time-sensitive situations.¹⁹ LLMs may also facilitate time-consuming tasks for IO practitioners, such as writing radiology reports.²⁰ Although LLMs have demonstrated promise, issues such as output hallucinations and source fabrication limit their present reliability.^{17,21} To provide responses to user prompts, LLMs may “hallucinate” and fabricate facts or reference nonexistent sources.^{22,23} Such hallucinations could be harmful, if not fatal, if patients or doctors were to assume incomplete/incorrect information to be trustworthy. Several methods have been developed to combat hallucinations; however, implementation varies widely and concerns about LLM bias and robustness remain.^{23,24} Limited model reliability may subsequently increase mistrust of LLMs, as has been publicized in previous occurrences.^{25,26} LLM models must be verified and validated aggressively to demonstrate sufficient reliability for health care integration.

CONSIDERATIONS FOR AI IMPLEMENTATION

Interpretability

Improving model interpretability is integral to adopting AI in health care. A significant barrier to integration is lack of model transparency, as quantitative features are difficult for humans to interpret. As such, models should be accompanied by explanations of develop-

ment and decision-making processes. Radiologists also require training to understand how models work and how to interpret findings. Technologic approaches to improving interpretability exist, such as feature attribution strategies that highlight imaging regions influential to model decision-making.²⁷ Prioritizing explainability will allow clinicians to make informed decisions, especially in cases where there is disagreement with AI.²⁸ To develop meaningful AI solutions that align with current IO approaches, interventional oncologists will need to play a significant role in developing and validating technology. Proper understanding of AI's logic and capabilities is necessary for widespread acceptance, as well as meaningful integration into patient-centered care.

Standardization

Prior to real-world implementation, AI models require standardization and testing to be trusted as a decision-making tool. They must be subjected to standard approval and regulatory processes to prevent negative impacts on patient care.

In the process of model development, it is ideal to standardize as many steps as possible to ensure reproducibility of findings. For example, many existing radiomics studies use standardized feature extraction algorithms, such as tools provided within the PyRadiomics package.²⁹ The widespread adoption of well-documented tool sets and algorithms help to standardize AI models and improve their performance. In addition, generalizability should be considered throughout model development. Although alterations to model structures and tuning of hyperparameters may help attain higher performance, they may limit generalizability in external testing and other applications. A balance must be found between continuous optimization and timely practical implementation.

Institutional variations in procedural standards may pose significant limitations to standardization, as image acquisition protocols and available imaging modalities may drastically shape model development. In addition, different administrative practices and data management systems may impact implementation across institutions. Approaches such as data augmentation and dropout attempt to account for institutional differences by increasing model versatility.³⁰ Future studies should explore additional approaches for broad applicability to ensure even integration across different institutions.

Furthermore, investigators should be expected to document the model development process to enable model acceptance. Standardized guidelines such as TRIPOD-AI (Transparent Reporting of a multivariable prediction model of Individual Prognosis Or

Diagnosis), PROBAST-AI (Prediction model Risk Of Bias ASsessment Tool), CLAIM (Checklist for Artificial Intelligence in Medical Imaging), and RQS (Radiomics Quality Score) should be employed when reporting imaging AI studies.^{31,32} Despite the increasing number of publications reporting novel AI applications, relatively few reference reporting quality guidelines. Publishers should establish such guidelines as expectations moving forward.

In 2021, the FDA issued the “Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.”^{33,34} The plan outlined five intended actions: (1) tailored regulatory framework, (2) good machine learning practice, (3) patient-centered approach incorporating transparency to users, (4) regulatory science methods regarding algorithm bias and robustness, and (5) real-world performance. Federal agencies such as the National Institute of Standards and Technology and the National Science Foundation have taken actions to promote knowledge, leadership, and coordination in establishing AI standards. International committees have also been established to define technical, clinical, and regulatory standards to ensure consistent operability of AI across clinical centers and geographic regions. Researchers and clinicians in IO will benefit from field-wide awareness and understanding of such efforts.

Model Assessment

A recurring theme within AI integration is the need to mitigate algorithm bias and improve robustness. Models must be subjected to thorough examination and validation by external parties to account for possible bias in development and reporting. Algorithms trained on imbalanced data may promote bias and exacerbate existing health disparities. However, data set descriptions are often lacking in information needed to assess such bias. In addition, high performance of AI models may be due to overfitting or confounding, leading to diminished performance upon external testing.^{35,36} As such, extensive external validation is necessary to assess model generalizability.

In addition, AI metrics rarely represent clinical applicability.³⁷ Before clinical implementation is considered, randomized controlled trials (RCTs) are necessary to evaluate an intervention’s impact. The SPIRIT-AI and CONSORT-AI extensions were established in 2020 to provide protocol and reporting guidelines, respectively, for clinical trials evaluating AI-related interventions, an important step in promoting transparency and rigor in AI research.^{38,39} However, there have been few RCTs conducted for AI technologies, with even fewer in IO.^{40,41}

Future RCTs should also consider evaluating performance of AI models when combined with humans to more realistically model clinical scenarios.^{42,43}

Clinical Implementation

To facilitate clinical adoption, AI models must be integrated into the radiologic workflow in a user-friendly manner. This will require the efforts of interventional oncologists familiar with AI to ensure usability. Clinicians must be able to interface with and understand such technology, and there must be mechanisms for physicians to evaluate AI technology and provide feedback, as well as opt out of integrating AI into workflows.

Before incorporating commercially available AI algorithms into clinical practice, institutions should test models on local data sets to ascertain suitability for their patient population. Depending on the intended generalizability of models, site-specific training may be necessary to adapt systems for local use. Furthermore, institutions should establish data registries and guidelines for monitoring model performance in clinical workflows. Such systems will allow institutions to evaluate the impact of AI models, as well as help identify potential areas for improvement and any potential safety concerns.

Models may require updates and eventual retraining, for which regulatory protocols should be established beforehand.⁴⁴ Although improvements are constantly being made in AI model performance, frequent implementation may result in drift. Updates should be limited and accompanied by comprehensive evaluation of clinical significance. It is also important to recognize that clinical and operational practices evolve over time, as do patient populations.⁴⁵ The introduction of novel algorithms may cause significant changes in practice, which will subsequently impact input data. Thus, it is necessary to carefully evaluate longitudinal performance and establish methods for identifying and addressing potential drift.

Expanding Data Sets, Multi-Institutional Efforts

A longstanding obstacle to the development and implementation of AI technology is the availability of sufficient, high-quality, and representative data, which is necessary to prevent algorithmic bias and improve model robustness.⁴⁶ The availability of such data would also increase feasibility of external testing, making it a justifiable expectation for model development. Data must be representative of the target population and unbiased to ensure model safety.

Most existing models were developed using single-institution data due to the lack of broad data sets. Most

health care data are not readily available for AI applications, contained within medical imaging archiving systems, electronic health records, and other systems that are difficult to consolidate. Data encoding is often inconsistent and requires great effort to standardize. These factors make it difficult to establish data sets of meaningful size. To increase data set sizes and population heterogeneity, highly organized collaborative efforts between multiple institutions are necessary to combine and curate comprehensive data sets. Developing larger data sets across institutional borders can bring IO closer to having generalizable models as opposed to individual institutions developing unique models. Specifically, a tiered approach where cross-institutional data consolidation is first performed on a city-wide basis may allow for gradual development of databases to reflect representative populations. This requires greater collaboration between institutions, while enforcing the same level of data protection. A potential solution for concerns about data sharing is federated learning, where patient data do not leave each hospital. This was demonstrated by PriMIA, an end-to-end method for medical imaging deep learning across multiple institutions.⁴⁷ Such tools may enable model development across institutions without patient data leaving hospital systems.

CONCLUSION

At present, there is a wealth of literature exploring the potential applications of AI to IO, with many displaying promising results. However, the translation of these efforts to clinical practice is often unclear. The future of AI within IO is dependent upon the development of infrastructure for standardized and responsible clinical implementation. Achieving this will require greater collaborative efforts within the field and frequent evaluation of long-term trajectory. ■

1. Letzen B, Wang CJ, Chapiro J. The role of artificial intelligence in interventional oncology: a primer. *J Vasc Interv Radiol*. 2019;30:38–41.e1. doi: 10.1016/j.jvir.2018.08.032
2. D'Amore B, Smolinski-Zhao S, Daye D, Uppot RN. Role of machine learning and artificial intelligence in interventional oncology. *Curr Oncol Rep*. 2021;23:70. doi: 10.1007/s11912-021-01054-6
3. Posa A, Barbieri P, Mazza G, et al. Technological advancements in interventional oncology. *Diagnostics (Basel)*. 2023;13:228. doi: 10.3390/diagnostics13020228
4. Leslie D, Mazumder A, Peppin A, et al. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*. 2021;372:n304. doi: 10.1136/bmj.n304
5. Morshid A, Elsayes KM, Khalaf AM, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell*. 2019;1:e180021. doi: 10.1148/ryai.2019180021
6. Luo Y-H, Xi IL, Wang R, et al. Deep learning based on MR imaging for predicting outcome of uterine fibroid embolization. *J Vasc Interv Radiol*. 2020;31:1010–1017.e3. doi: 10.1016/j.jvir.2019.11.032
7. Wesdorp NJ, Hellingman T, Jansma EP, et al. Advanced analytics and artificial intelligence in gastrointestinal cancer: a systematic review of radiomics predicting response to treatment. *Eur J Nucl Med Mol Imaging*. 2021;48:1785–1794. doi: 10.1007/s00259-020-05142-w
8. Liu D, Liu F, Xie X, et al. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol*. 2020;30:2365–2376. doi: 10.1007/s00330-019-06553-6

9. Daye D, Staziaki PV, Furtado VF, et al. CT texture analysis and machine learning improve post-ablation prognostication in patients with adrenal metastases: a proof of concept. *Cardiovasc Intervent Radiol*. 2019;42:1771–1776. doi: 10.1007/s00270-019-02336-0
10. Liu Y, Chen X, Wang Z, et al. Deep learning for pixel-level image fusion: recent advances and future prospects. *Information Fusion*. 2018;42:158–173. doi: 10.1016/j.inffus.2017.10.007
11. Newbury A, Ferguson C, Valero DA, et al. Interventional oncology update. *Eur J Radiol Open*. 2022;9:100430. doi: 10.1016/j.ejro.2022.100430
12. Li X, Young AS, Raman SS, et al. Automatic needle tracking using Mask R-CNN for MRI-guided percutaneous interventions. *Int J Comput Assist Radiol Surg*. 2020;15:1673–1684. doi: 10.1007/s11548-020-02226-8
13. Mwikirize C, Noshier JL, Hachililoglu I. Convolution neural networks for real-time needle detection and localization in 2D ultrasound. *Int J Comput Assist Radiol Surg*. 2018;13:647–657. doi: 10.1007/s11548-018-1721-y
14. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. *Neurosurgery*. Published online June 12, 2023. doi: 10.1227/neu.0000000000002551
15. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. *arXiv:2305.09617*. May 16, 2023. Accessed August 8, 2023. <https://arxiv.org/pdf/2305.09617>
16. Nori H, King N, McKinney SM, et al. Capabilities of GPT-4 on medical challenge problems. Published March 20, 2023. Updated April 12, 2023. Accessed August 8, 2023. <https://arxiv.org/pdf/2303.13375>
17. Rahsepar AA, Tavakoli N, Kim GHJ, et al. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology*. 2023;307:e230922. doi: 10.1148/radiol.230922
18. Sorin V, Klang E, Sklar-Levy M, et al. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer*. 2023;9:44. doi: 10.1038/s41523-023-00557-8
19. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2:e0000198. doi: 10.1371/journal.pdig.0000198
20. Jeblick K, Schachtner B, Dext J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. December 30, 2022. Accessed August 8, 2023. <https://arxiv.org/pdf/2212.14882>
21. Bang Y, Cahyawijaya S, Lee N, et al. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. Published February 8, 2023. Updated February 28, 2023. Accessed August 8, 2023. <https://arxiv.org/pdf/2302.04023>
22. Feldman P, Foulds JR, Pan S. Trapping LLM "hallucinations" using tagged context prompts. June 9, 2023. Accessed August 29, 2023. <https://arxiv.org/pdf/2306.06085>
23. Manakul P, Liusie A, Gales MJ. Selfcheckgpt: zero-resource black-box hallucination detection for generative large language models. Published March 15, 2023. Updated May 8, 2023. Accessed August 29, 2023. <https://arxiv.org/pdf/2303.08896>
24. Zhuo TY, Huang Y, Chen C, Xing Z. Red teaming ChatGPT via jailbreaking: bias, robustness, reliability and toxicity. Published January 30, 2023. Updated May 29, 2023. Accessed August 29, 2023. <https://arxiv.org/pdf/2301.12867>
25. Weiser B. Here's what happens when your lawyer uses ChatGPT. *The New York Times*. May 27, 2023. Accessed July 7, 2023. <https://www.nytimes.com/2023/05/27/nyregion/aviana-airline-lawsuit-chatgpt.html>
26. Alkasssi H, McFarlane SL. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;15:e35179. doi: 10.7759/cureus.35179
27. Reyes M, Meier R, Pereira S, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities. *Radiol Artif Intell*. 2020;2:e190043. doi: 10.1148/ryai.2020190043
28. Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20:310. doi: 10.1186/s12911-020-01332-6
29. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107. doi: 10.1158/0008-5472.Can-17-0339
30. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
31. Collins GS, Dhiman P, Navarro CLA, 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. doi: 10.1136/bmjopen-2020-048008
32. Mongan J, Moy L, Kahn Jr CE. Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2:e200029. doi: 10.1148/ryai.2020200029
33. US Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). Accessed August 29, 2023. https://downloads.regulations.gov/FDA-2019-N-1185-0068/attachment_2.pdf
34. US Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. January 2021. Accessed August 29, 2023. <https://www.fda.gov/media/145022/download>
35. England JR, Cheng PM. Artificial intelligence for medical image analysis: a guide for authors and reviewers. *AJR Am J Roentgenol*. 2019;212:513–519. doi: 10.2214/ajr.18.20490
36. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell*. 2022;4:e210064. doi: 10.1148/ryai.210064
37. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med*. 2018;1:40. doi: 10.1038/s41746-018-0048-y

38. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26:1364-1374. doi: 10.1038/s41591-020-1034-x
39. Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med.* 2020;26:1351-1363. doi: 10.1038/s41591-020-1037-7
40. Plana D, Shung DL, Grimshaw AA, et al. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw Open.* 2022;5:e2233946-e2233946. doi: 10.1001/jamanet-workopen.2022.33946
41. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ.* 2020;368:m689. doi: 10.1136/bmj.m689
42. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med.* 2022;28:31-38. doi: 10.1038/s41591-021-01614-0
43. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25:44-56. doi: 10.1038/s41591-018-0300-7
44. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195. doi: 10.1186/s12916-019-1426-2
45. Nestor B, McDermott MBA, Chauhan G, et al. Rethinking clinical prediction: why machine learning must consider year of care and feature aggregation. November 30, 2018. Accessed August 29, 2023. <https://arxiv.org/pdf/1811.12583>
46. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology.* 2020;295:4-15. doi: 10.1148/radiol.2020192224
47. Kaissis G, Ziller A, Passerat-Palmbach J, et al. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat Med Intell.* 2021;3:473-484. doi: 10.1038/s42256-021-00337-8

CO-FIRST AUTHOR**Helen Zhang, BS**

Department of Diagnostic Radiology
Rhode Island Hospital
Warren Alpert Medical School of Brown University
Providence, Rhode Island
Disclosures: None.

CO-FIRST AUTHOR**Shreyas Kulkarni, BS**

Department of Diagnostic Radiology
Rhode Island Hospital
Warren Alpert Medical School of Brown University
Providence, Rhode Island
Disclosures: None.

Zhicheng Jiao, PhD

Department of Diagnostic Radiology
Rhode Island Hospital
Warren Alpert Medical School of Brown University
Providence, Rhode Island
Disclosures: None.

Harrison X. Bai, MD, MS

Department of Radiology and Radiological Sciences
Johns Hopkins University School of Medicine
Baltimore, Maryland
hbai7@jhu.edu
Disclosures: None.