ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

EDITORIAL

Artificial intelligence as an adjunct tool for breast oncologists — are we there yet?

Over the past few years, there has been a renewed and exponential public and scientific interest in the breadth of applicability of artificial intelligence (AI) in medicine. This has prompted a rise in dedicated medical journals and podcasts, such as the recent announcement of the 2024 launch of NEJM AI (ai.nejm.org), focused solely on distilling high-quality evidence and reviewing the potential of medical AI.

AI involves computer systems that can mimic intelligent human behavior in interpreting information and performing specific tasks. Machine learning (ML) is a branch of AI that uses statistical methods to optimize models for a specific task. In this editorial, we review the utility, promise, and pitfalls of AI integration into our daily practice as breast medical oncologists, highlighting some specific clinical and academic implications (summarized in Figure 1).

AI IN CLINICAL PRACTICE

There are multiple potential applications of AI technology along the cancer care continuum of patients with breast cancer, spanning from early detection to diagnosis to treatment to survivorship. AI can be used to develop testing tools to provide more accurate results with the goal to improve patient outcomes. For example, in breast cancer screening, which has had a profound impact on breast cancer outcomes, AI models have been used to aid radiologists in reading mammograms with increased accuracy.[1] Similarly, breast ultrasounds can be read with increased specificity, without impacting the sensitivity.[2] AI can also have an impact within pathology, which is critical to establish a breast cancer diagnosis and to determine the optimal treatment for an individual patient.[3] In one study, the inter-observer diagnostic concordance for breast biopsy diagnoses among pathologists was only 75%, with the most variability noted among ductal carcinoma *in situ* and atypia diagnoses.[4] AI-based algorithms have been proposed to improve reproducibility and accuracy.[5] Another example is the testing of the human epidermal growth receptor factor 2 (HER2), for which there is a high rate of discordance between pathologists. With the development of novel therapeutic agents that require higher sensitivity, HER2 testing potentially could be improved using AI models to aid personalized medicine.[6-8]

Models are also being developed for breast cancer risk assessment and to guide treatment decisions. For example, being able to predict which patients are more likely to achieve a pathologic complete response with preoperative therapy will allow clinicians to tailor treatments, intensifying therapies for those who are not likely to respond to standard treatment and de-escalating for those with very high likelihood to respond, to improve patient outcomes and decrease risk for toxicities.[9,10] One example is an Image-based Risk Score (IbRiS) developed by one group utilizing an ML algorithm with an 84.15% accuracy in classifying estrogen receptor-positive breast cancer into low- and high-risk scores, offering a cheaper alternative to genomic assays.[11] To date, most of these tools are in the experimental phase, but we anticipate that some models may be implemented in clinical practice soon.

AI could help in decision making of physicians; however, given that some natural language processing (NLP) models may be outdated, all the information needs to be verified before making any medical decisions that could impact patient care. For example, ChatGPT has a knowledge cut-off from September 2021, and significant advances in oncology have occurred since; therefore if we were to ask it a clinical question for which newer information is available, the answer would likely be incorrect or outdated. In addition, given that these models use data from the web, not always verified or peer reviewed, there is a high risk for errors and to introduce bias if used in clinical practice. For example, Watson for Oncology, launched by IBM (New York, NY), applied AI algorithms to formulate treatment plans for patients with breast cancer, with one study showing similar suggestions when compared with a multidisciplinary tumor board.[12] However, another recent study by Lukac et al.[13] utilizing ChatGPT as a supportive tool for multidisciplinary tumor board in patients with early breast cancer noted that neoadjuvant treatments were not considered by ChatGPT, HER2 expression interpretation was incorrect, and, as expected, recent studies were not incorporated in the decision recommendations.

NLP models such as ChatGPT (OpenAI, San Francisco, CA) and Bidirectional Encoder Representations from Transformers (BERT) could help reduce the documentation burden that health care providers often face. The role of these models in creating visit reports for patients and discharge summaries as clinical documentations is actively being studied.[14] NLP models could be trained to obtain intake forms, to write notes, and to aid with other documentation, such as prior authorizations and letters.[15] They

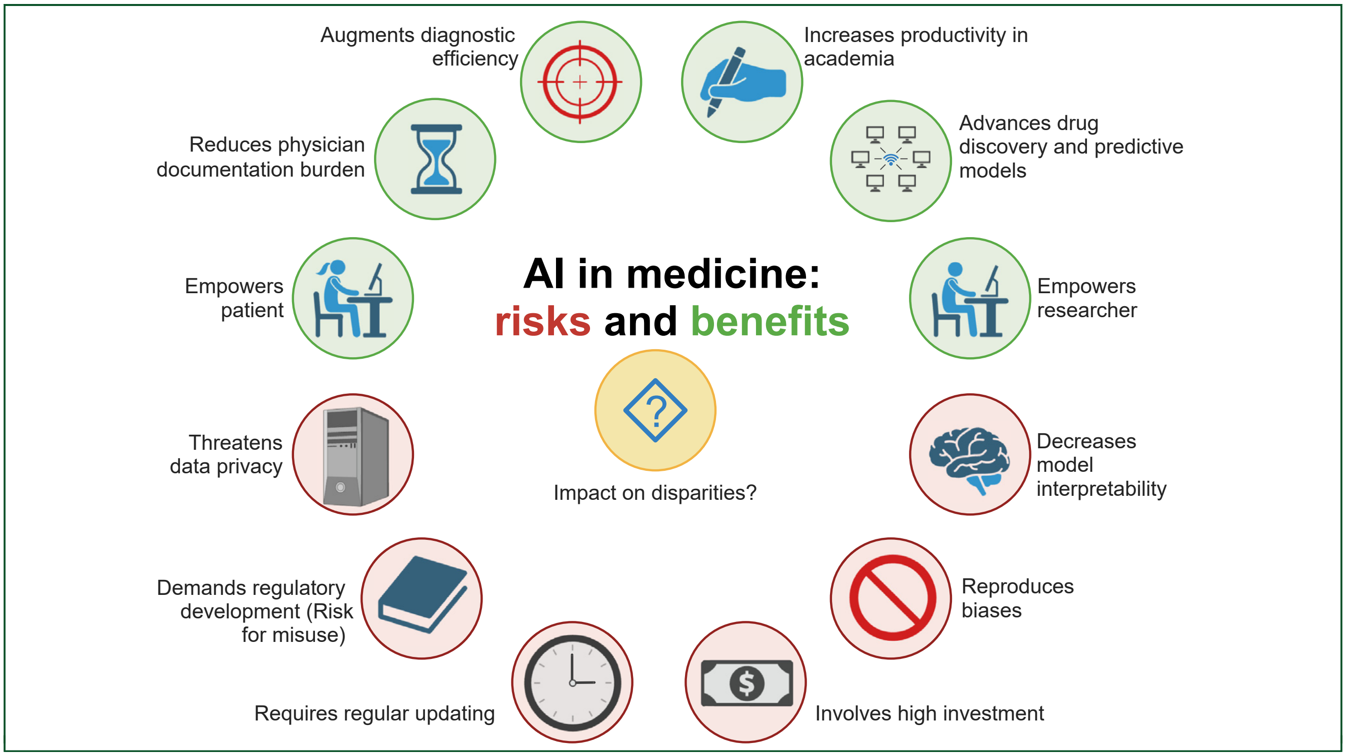

Figure 1. Utility, promise, and pitfalls of artificial intelligence (AI) in oncology. Created with BioRender.

can also function as virtual assistants, for example, scheduling meeting and reviewing emails, to decrease the administrative burden that many clinicians face. This could have a significant impact on providers well-being.[16] At every stage of a clinical trial, AI holds much utility in streamlining the process with enhancement of protocol development, efficient patient-trial matching, and real-time data collection from patients.[1] However, OpenAI models, such as ChatGPT, are not Health Insurance Portability and Accountability Act (HIPAA) compliant and caution is needed, as patient-identifiable information should not be entered in these open publicly available platforms. Several HIPAA-compliant models are being developed. Once ready for use in clinical practice, ideally these models will be widely accessible as this could help provide patients with less resources with better quality health care and provide care in areas with limited resources, with the potential to narrow health care disparities worldwide. As with any new technology, it is important to rigorously validate new tools prior to deploying them in marginalized populations already on the fringes of health care.

AI FROM THE PATIENT PERSPECTIVE

The use of a search of the internet to try to answer medical questions is not a new practice.[17] With the increased access to the internet and smartphones, patients have access to large amounts of information, which depending on the accuracy of the sources can help patients in decision making. It is important to mention that many people in the United States and worldwide do not have access to these technologies, and are at an additional disadvantage. The use of NLPs for decision making is also a double-edged sword as it can summarize large amounts of information and can simplify information for patients with limited health care literacy to improve their understanding, but NLPs are prone to unpredictable errors and can represent a source of misinformation.[15]

AI could also be a valuable tool for translation. NLPs have the ability to translate to multiple languages. This could be used to provide patients with information about treatments, appointments, prescriptions, and much more, helping overcome health care disparities.

AI IN RESEARCH

The potential for AI in research is very broad. The current academic environment demands productivity, often measured in grants and publications. The amount of time needed to write grant proposals, research protocols, and papers is very significant, and NLPs could be implemented to help with this. The potential benefit is that researchers could spend more time discussing new ideas, developing new agents, and/or seeing patients. This could also be a tool to aid writers with limited English proficiency.[15] There are several AI tools that can be used to create images for medical research. There are, however, significant limitations for this: (i) All the output from NLPs needs to be verified and confirmed, as NLPs often make subtle mistakes (known as hallucinations). NLP needs to be validated to start integrating this into clinical practice and research[13]; (ii) The role of NLPs in authorship needs to be regulated and best practice guidelines are needed; although there are programs available to detect AI-generated content, there is a critical need to determine rules for AI use in this setting[13,18,19]; (iii) NLPs are prone to bias[13]; and (iv) There is a risk for plagiarism.

AI also has potential in the research process. For example, NLPs can write code that could be used to create research models or to collect data from the electronic medical records to create databases that can be updated periodically.[20] ML models are being used broadly in research, including in prediction models, improving pathology, and radiology reports, as mentioned earlier.

CONCLUSIONS

To address the challenges and maximize the promises of AI technology in our daily workflow as medical oncologists, best practice guidelines need to be developed by medical societies and academic institutions. Areas of need include how to incorporate AI in clinical research, how to determine authorship, and how to assess productivity in research that incorporates AI technology. Lastly, and before incorporating AI in clinical practice, best practices must be developed for assessing the accuracy and reliability of AI algorithms in clinical settings. This can include regular audits and reviews to ensure that the algorithms are providing accurate results and that any errors are identified and corrected promptly.

In conclusion, AI technology, if used appropriately, has the potential to revolutionize academic medicine and improve the oncological care of our patients. Developing best practices that prioritize transparency, reproducibility, and accuracy is essential.

I. Schlam[1,2]*, M. C. Saad Menezes[2,3], C. Corti[4,5], A. Tan[3], I. Abuali[6] & S. M. Tolaney[7,8,9]

[1]*Department of Hematology and Oncology, Tufts Medical Center, Boston;*
[2]*Harvard T.H. Chan School of Public Health, Boston;*
[3]*Department of Biomedical Informatics, Harvard Medical School, Boston, USA;*
[4]*Division of New Drugs and Early Drug Development for Innovative Therapies, European Institute of Oncology, IRCCS, Milan;*
[5]*Department of Oncology and Hemato-Oncology (DIPO), University of Milan, Milan, Italy;*
[6]*Department of Hematology and Oncology, Massachusetts General Hospital, Boston;*
[7]*Department of Medical Oncology, Dana-Farber Cancer Institute, Boston;*
[8]*Breast Oncology Program, Dana-Farber Brigham Cancer Center, Boston;*
[9]*Harvard Medical School, Boston, USA*
(*E-mail: ilana.schlam@tuftsmedicine.org).

Available online xxx

https://doi.org/10.1016/j.esmoop.2023.101643

FUNDING

None declared.

DISCLOSURE

SMT reports consulting or advisory role for Novartis, Pfizer, Merck, Eli Lilly, AstraZeneca, Genentech/Roche, Eisai, Sanofi, Bristol Myers Squibb, Seattle Genetics, CytomX Therapeutics, Daiichi-Sankyo, Gilead, Ellipses Pharma, 4D Pharma, OncoSec Medical Inc., BeyondSpring Pharmaceuticals, OncXerna, Zymeworks, Zentalis, Blueprint Medicines, Reveal Genomics, ARC Therapeutics, Infinity Therapeutics, Myovant, Zetagen, Umoja Biopharma, Artios Pharma, Menarini/Stemline, Aadi Biopharma, Bayer, Incyte Corp, and Jazz Pharmaceutical; and research funding from Genentech/Roche, Merck, Exelixis, Pfizer, Lilly, Novartis, Bristol Myers Squibb, Eisai, AstraZeneca, Gilead, NanoString Technologies, Seattle Genetics, and OncoPep. All other authors have declared no conflicts of interest.

REFERENCES

1. Cesario A, Simone I, Paris I, et al. Development of a digital research assistant for the management of patients' enrollment in oncology clinical trials within a research hospital. *J Pers Med*. 2021;11(4):244.
2. Shen Y, Shamout FE, Oliver JR, et al. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. *Nat Commun*. 2021;12(1):5645.
3. Ibrahim A, Gamble P, Jaroensri R, et al. Artificial intelligence in digital breast pathology: techniques and applications. *Breast*. 2020;49:267-273.
4. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *J Am Med Assoc*. 2015;313(11):1122-1132.
5. Weaver DL. Pathology evaluation of sentinel lymph nodes in breast cancer: protocol recommendations and rationale. *Mod Pathol*. 2010;23(suppl 2):S26-S32.
6. Robbins CJ, Fernandez AI, Han G, et al. Multi-institutional assessment of pathologist scoring HER2 immunohistochemistry. *Mod Pathol*. 2023;36(1):100032.
7. Palm C, Connolly CE, Masser R, et al. Determining HER2 status by artificial intelligence: an investigation of primary, metastatic, and HER2 low breast tumors. *Diagnostics (Basel)*. 2023;13(1):168.
8. Cho SI, Choi S, Lee T, et al. Exploring expression levels of HER2, HER3, MET, Claudin18.2, and MUC16 across 16 cancer types using an artificial intelligence-powered immunohistochemistry analyzer [abstract]. *J Clin Oncol*. 2023;41(suppl 16):3135.
9. Garufi G, Maliziola N, Monte ED, et al. Artificial intelligence (AI)—based machine learning models (ML) for predicting pathological complete response (pCR) in patients with hormone receptor (HoR)—positive/HER2-negative early breast cancer (EBC) undergoing neoadjuvant chemotherapy (NCT): a retrospective cohort study [abstract]. *J Clin Oncol*. 2023;41(suppl 16):597.
10. Liu VY, Selvaraju RR, Chen E, et al. Development and validation of an AI-derived digital pathology-based biomarker to predict benefit of long-term androgen deprivation therapy with radiotherapy in men with localized high-risk prostate cancer across multiple phase III NRG/RTOG trials [abstract]. *J Clin Oncol*. 2023;41(suppl 16):5001.
11. Basavanhally A, Jun X, Madabhushi A, Ganesan S. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with Oncotype DX assay. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. Boston, MA: IEEE; 2009:251-254.
12. Somashekhar SP, Sepúlveda MJ, Puglielli S, et al. Watson for Oncology and breast cancer treatment recommendations: agreement

with an expert multidisciplinary tumor board. *Ann Oncol*. 2018;29(2):418-423.

13. Lukac S, Dayan D, Fink V, et al. Evaluating ChatGPT as an adjunct for the multidisciplinary tumor board decision-making in primary breast cancer cases. *Arch Gynecol Obstet*. 2023. https://doi.org/10.1007/s00404-023-07130-5. In press.
14. Lilly CM, Cucchi E, Marshall N, Katz A. Battling intensivist burnout: a role for workload management. *Chest*. 2019;156(5):1001-1007.
15. Lee P, Goldberg C, Kohane Z. *The AI Revolution in Medicine GPT-4 and Beyond*. New York, New York: Pearson; 2023.
16. Kundu S. How will artificial intelligence change medical training? *Commun Med (Lond)*. 2021;1(1):8.
17. Cacciamani GE, Dell'Oglio P, Cocci A, et al. Asking "Dr. Google" for a second opinion: the devil is in the details. *Eur Urol Focus*. 2021;7(2):479-481.
18. Else H. Abstracts written by ChatGPT fool scientists. *Nature*. 2023;613(7944):423.
19. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379(6630):313.
20. Lee S, Kim HS. Prospect of artificial intelligence based on electronic medical record. *J Lipid Atheroscler*. 2021;10(3):282-290.