

## EDITORIAL OPEN



# A framework for artificial intelligence in cancer research and precision oncology

npj Precision Oncology (2023)7:43; <https://doi.org/10.1038/s41698-023-00383-y>

## INTRODUCTION

This collection aims to disseminate the latest research and advancements in all aspects of artificial intelligence (AI) within the realm of cancer research, including basic, translational, and clinical research. The collection is undoubtedly ambitious, as it seeks to present a comprehensive review of current AI applications in precision oncology while offering expert insights on accelerating the transition of AI tools from the laboratory to the clinic, with the aim of ultimately enhancing patient care and improving clinical endpoints. Priority will be given to articles that employ innovative methodologies, address pertinent real-world issues, and provide robust evidence through the utilization of multicentric datasets. Although the definition of AI remains ambiguous, most contemporary, successful AI systems incorporate some form of deep learning<sup>1</sup>. Within the biomedical research literature, there is a prevailing consensus that the term “AI” can be applied to deep learning approaches—those involving models with millions or billions of parameters—rather than simple statistical models such as linear regression with only a few free parameters and co-variables<sup>1</sup>. Although deep learning approaches are the primary focus of this collection, we remain receptive to related fields.

## IMAGE PROCESSING

Medicine and oncology, in particular, are abundant with image data<sup>2,3</sup>. For example, skin tumors are diagnosed visually and can be identified from digital photographs. Numerous cancer types, such as those of the digestive tract, are diagnosed endoscopically. Breast cancer, and many other tumor types, are screened for and detected by radiology imaging. The definitive diagnosis of cancer is made through a process of histopathology based on the analysis of images from stained tissue. Radiology images are also commonly used in the staging and evaluation of responses to cancer treatments. In essence, the heavy reliance of oncology on image analysis makes the acquisition, interpretation, and analysis of such images critical at all stages of the cancer detection, characterization, and follow-up processes. Over the last century, a critical revolution has taken place in the field of medical imaging processing and storage, transitioning from analog detectors in radiology to digital detectors. Additionally, there has been an extensive expansion in the efforts of pathology departments in the digitalization of tumor samples. This shift toward digitalization has opened up new and exciting opportunities for the storage, sharing, and use of image data. With digital image data, powerful computational AI methods can be applied, enabling high-throughput analysis and exploitation of data that was once only subjectively assessed by medical experts. As a result, the digitization of medical imaging represents a major breakthrough and promises significant advances in oncology. Prior to 2012, computer-based image analysis was a formidable challenge, and

computer-based methods were highly restricted to very narrow domains and datasets. With the breakthrough of AlexNet<sup>4</sup> as the first prominent convolutional neural network (CNN) in 2012, deep learning has since become the method of choice for virtually any image analysis task. Consequently, in oncology, the vast majority of AI solutions are related to the analysis of digital images. AI can automate human tasks, such as cancer detection on endoscopy videos, radiology images, or histopathology slides. Numerous programs have received regulatory approval in Europe, the US, and other markets and can now be utilized in patients’ routine care. AI can, however, transcend these applications by yielding image-based biomarkers capable of directly predicting survival or treatment response, thus extracting considerably more information from image data than was previously possible<sup>5,6</sup>. Such biomarkers could be integrated into prognostic or predictive models that are already utilized in clinical care, thus acting as integrative biomarkers. In this collection, we aim to explore all these approaches spanning cancer diagnosis, treatment, and follow-up.

## LANGUAGE PROCESSING

The text serves as the universal interface for communication among humans and between humans and machines. Electronic health records, patient-reported outcomes, and communication between healthcare providers and patients all rely on text. In oncology, this implies that a significant portion of diagnostic and clinical workflows generates text artifacts<sup>7</sup>. Physicians often devote a considerable amount of time to reading text documents describing patients’ previous medical histories or diagnostic reports. Until recently, this vast resource remained largely untapped by digital tools. The field of natural language processing (NLP) has developed numerous methods for computer-based text mining, but none were particularly robust, effective, and generalizable. Tasks such as summarizing multi-page medical reports, translating languages, or converting writing styles were considered highly complex. With the advent of transformer neural networks, initiated by the groundbreaking paper “Attention is All You Need”<sup>8</sup>, exponential growth in NLP has been spurred. By employing transformers with tens or hundreds of billions of parameters (known as large language models, or LLMs) and training them on vast amounts of data (primarily all available internet text), human expert-level performance can be achieved for nearly any language understanding or creation task. In an unexpected twist of history, LLMs have also inadvertently demonstrated remarkable logical reasoning skills<sup>9</sup>. The transformative impact of this technology on all facets of society, including cancer research and oncology, cannot be overstated. The introduction of LLMs will likely lead to an exponential increase in overall productivity and pace in the field, much like the introduction of the Internet. This does not imply that cancer researchers or oncologists will be rendered obsolete; rather, the demand for human orchestration in the process will continue. In this collection, we encourage the submission of scientifically rigorous articles that employ LLM-based methods in oncology and provide empirical data supporting their benefits for research and patient care.

## ENHANCING GENOMIC INFORMATION WITH AI

The use of next-generation sequencing (NGS) has revolutionized cancer research and drug development in oncology, leading to unparalleled precision and success. AI tools open tremendous possibilities to leverage high-throughput genomics data, including whole-genome, whole-exome, or whole-transcriptome analysis, with the promise not only to enhance the clinical practice but also to acquire valuable insights into cancer biology<sup>10</sup>. The implementation of AI-based techniques for the analysis of genomics data can identify novel genomic and transcriptomic features with prognostic and predictive significance in cancer, as well as enable large-scale analysis of pathogenicity of variants and genotype–phenotype associations. Importantly, AI applications in cancer genomics could unlock a new level of decision-support tools enabling truly personalized care. There remain several challenges that need to be addressed in the adoption of AI analytics in genomics. With this in mind, we encourage the submission of articles detailing clinically meaningful discoveries in genomics using AI, as well as new methods for tackling encountered challenges.

## THE FUTURE IS MULTIMODAL

The tremendous success of transformer neural networks has brought another once-elusive goal within reach: multimodal machine learning models<sup>11</sup>. These models can accept different types of data as a simultaneous input, such as various image types, images combined with text, or images integrated with genomic data. Intuitively, it is clear that oncological decision-making is inherently multimodal. Oncologists do not prescribe treatments based solely on a single image; rather, they consider multiple pieces of information and corresponding data. Combining different data types can yield more than just the sum of their parts. Not uncommonly, a pattern in an image only becomes meaningful in light of the genetic make-up of the patient, a given cancer genomic alteration, and/or medical history. Training machine learning models on multimodal data was once considered too ambitious. This changed recently, as several publications in 2022 and early 2023 demonstrated sensible applications of multimodal models in oncology<sup>12,13</sup>. By utilizing transformer neural networks, these approaches have become more powerful and versatile<sup>14</sup>. While such approaches present regulatory and practical challenges, we believe that their implementation is imminent. In this collection, we encourage the submission of articles exploring novel technical approaches for multimodality in relevant oncological applications.

## THE PRIMACY OF EVIDENCE-BASED MEDICINE AND PATIENT-CENTERED CARE

This collection aims to provide a comprehensive review of current AI applications in precision oncology and explore novel technical approaches. The overarching aim is the use of AI to generate solid scientific evidence which can improve cancer research and patient care. We currently live in an era of unprecedented technological progress, with no signs of slowing down<sup>15</sup>, and this collection aims to reflect these developments. However, we emphasize that the practice of medicine transcends technological advancement; at its core, it involves empathetic human interactions, shared decision-making based on rigorous science, and consideration of patient preferences. These principles must guide any research we undertake, making it imperative that AI in oncology respects and advances the primacy of evidence-based medicine and patient-centered care. We must be mindful of these principles, and in addition, we must include ethical considerations, data privacy concerns, and potential biases in the development, evaluation, and clinical implementation of AI systems.

## METHODS

In accordance with the COPE (Committee on Publication Ethics) position statement of 13 February 2023 (<https://publicationethics.org/cope-position-statements/ai-author>), the authors hereby disclose the use of the following artificial intelligence models during the writing of this article. GPT-4 (OpenAI) for checking spelling and grammar.

Published online: 17 May 2023

Raquel Perez-Lopez<sup>1,6</sup>, Jorge S. Reis-Filho<sup>2,6</sup> and Jakob Nikolas Kather<sup>3,4,5,6</sup>

<sup>1</sup>Radiomics Group, Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. <sup>2</sup>Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>3</sup>Else Kroener Fresenius Center for Digital Health, Technical University Dresden, Dresden, Germany. <sup>4</sup>Department of Medicine I, University Hospital Dresden, Dresden, Germany. <sup>5</sup>Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany. <sup>6</sup>These authors contributed equally: Raquel Perez-Lopez, Jorge S. Reis-Filho, Jakob Nikolas Kather. ✉email: rperez@vhio.net; reisfilj@mskcc.org; jakob-nikolas.kather@alumni.dkfz.de

## REFERENCES

- Shmatko, A., Ghaffari Laleh, N., Gerstung, M. & Kather, J. N. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat. Cancer* **3**, 1026–1038 (2022).
- Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).
- Bera, K., Braman, N., Gupta, A., Velcheti, V. & Madabhushi, A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat. Rev. Clin. Oncol.* **19**, 132–146 (2022).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).
- Reis-Filho, J. S. & Kather, J. N. Overcoming the challenges to implementation of artificial intelligence in pathology. *J. Natl. Cancer Inst.* <https://doi.org/10.1093/jnci/djad048> (2023).
- Ghaffari Laleh, N., Ligerio, M., Perez-Lopez, R. & Kather, J. N. Facts and hopes on the use of artificial intelligence for predictive immunotherapy biomarkers in cancer. *Clin. Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-22-0390> (2022).
- Rösler, W. et al. An overview and a roadmap for artificial intelligence in hematology and oncology. *J. Cancer Res. Clin. Oncol.* <https://doi.org/10.1007/s00432-023-04667-5> (2023).
- Vaswani, A. et al. Attention is all you need. Preprint at *arXiv* <http://arxiv.org/abs/1706.03762> (2017).
- Bubeck, S. et al. Sparks of artificial general intelligence: early experiments with GPT-4. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2303.12712>.
- Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W. & Mostafavi, S. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* **24**, 125–137 (2023).
- Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**, 1095–1110 (2022).
- Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2022).
- Foersch, S. et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat. Med.* <https://doi.org/10.1038/s41591-022-02134-1> (2023).
- Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
- Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).

## AUTHOR CONTRIBUTIONS

The authors jointly conceived and wrote this article.

## FUNDING

R.P.L. is supported by LaCaixa Foundation, a CRIS Foundation Talent Award (TALENT19-05), the FERO Foundation, the Instituto de Salud Carlos III-Investigación en Salud (PI18/01395 and PI21/01019), the Prostate Cancer Foundation (18YOUN19) and the Asociación Española Contra el Cáncer (AECC) (PRYCO211023SERR, funding C.M.). J.N.K. is supported by the German Federal Ministry of Health (DEEP LIVER, ZMV11-2520DAT111) and the Max-Eder-Program of the German Cancer Aid (grant #70113864), the German Federal Ministry of Education and Research (PEARL, 01KD2104C; CAMINO, 01EO2101; SWAG, 01KD2215A; TRANSFORM LIVER, 031L0312A; TANGERINE, 01KT2302 through ERA-NET Transcan), the German Academic Exchange Service (SECAI, 57616814), the German Federal Joint Committee (Transplant.KI, 01VSF21048) the European Union's Horizon Europe and innovation program (ODELIA, 101057091; GENIAL, 101096312) and the National Institute for Health and Care Research (NIHR, NIHR213331) Leeds Biomedical Research Centre. JSR-F is funded in part by the Breast Cancer Research Foundation, a Susan G Komen Leadership grant, an NIH/NCI P50 CA247749 01 grant, and an NIH/NCI grant No. P30CA008748. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

## COMPETING INTERESTS

R.P.L. reports grants from Roche and AstraZeneca and a consulting or advisory role at Roche. J.S.R.F. reports a leadership (board of directors) role at Grupo Oncoclinicas,

stock or other ownership interests at Repare Therapeutics and Paige.AI, and a consulting or advisory role at Genentech/Roche, Invicro, Ventana Medical Systems, Volition RX, Paige.AI, Goldman Sachs, Bain Capital, Novartis, Repare Therapeutics, Lilly, Saga Diagnostics, MultiplexDX, and Personalis. J.N.K. declares consulting services for Owkin, France; DoMore Diagnostics, Norway and Panakeia, UK and has received honoraria for lectures by Eisai, Roche, MSD, and Fresenius. The remaining authors declare no competing interests.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023