frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 22 March 2023
DOI 10.3389/fonc.2023.1184428

OPEN ACCESS

EDITED AND REVIEWED BY
Timothy James Kinsella,
Brown University, United States

*CORRESPONDENCE
Chunhao Wang
chunhao.wang@duke.edu

SPECIALTY SECTION
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

RECEIVED 11 March 2023
ACCEPTED 17 March 2023
PUBLISHED 22 March 2023

CITATION
Hrinivich WT, Wang T and Wang C (2023) Editorial: Interpretable and explainable machine learning models in oncology. *Front. Oncol.* 13:1184428. doi: 10.3389/fonc.2023.1184428

Editorial: Interpretable and explainable machine learning models in oncology

William Thomas Hrinivich[1], Tonghe Wang[2] and Chunhao Wang[3]*

[1]Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University, Baltimore, MD, United States, [2]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [3]Department of Radiation Oncology, Duke University, Durham, NC, United States

KEYWORDS

explainable machine learning, oncology, explainable deep learning, radiotherapy, MRI, outcome prediction, ultrasound

Editorial on the Research Topic
Interpretable and explainable machine learning models in oncology

The application of machine learning (ML) in cancer diagnosis and treatment is rapidly expanding, capitalizing on the availability of highly-detailed patient-specific data, advancements in hardware performance, and algorithmic breakthroughs. ML approaches have been applied to all facets of oncology (1) including medical image reconstruction (2) and classification (3), interpretation of histopathology (4, 5), genomic analysis (6), chemotherapy and radiotherapy outcome prediction (7; 8) and surgical guidance (9). As described by Lu et al. in this collection, model "explainability" may refer to the ability to describe elements of the model and "interpretability" may refer to the ability to understand reasoning behind the model's prediction. ML model explainability and interpretability (MEI) are of growing interest (10) as models grow in complexity through highly non-linear approaches such as deep learning (DL), which may offer unparalleled performance but provide little or no inherent MEI (11).

Limited MEI associated with sophisticated ML approaches creates concerns for investigators and clinicians (12). In situations when ML models are used to guide critical decision making for a patient's cancer care, poor MEI may introduce safety concerns preventing clinical adoption and degrading overall trust in model efficacy. Limited MEI may prevent human users from making informed decisions regarding the relevance of a model's output to a specific scenario or from supplementing a prediction with their existing knowledge. These issues are evidenced by instances where large models provided predictions based on extraneous features unrelated to patient clinicopathological or physiological features due to biases in training data (13). Finally, limited MEI poses challenges for investigators pursuing improved performance by hindering interpretation of factors impacting model accuracy and generalizability.

Accordingly, many approaches have been proposed to improve MEI in oncology and other domains through a variety of approaches that may be considered model-specific or model-agnostic, and *ad-hoc* or *post-hoc*. Lu et al. (11, 14, 15), Model-specific approaches

are only applicable for specific models, such as certain saliency map approaches for convolutional neural networks (16). Model-agnostic approaches may be applied more generally, often numerically investigating the relationships between model inputs and outputs (17). *Ad-hoc* methods include approaches intended to make the model intrinsically explainable and may include hand-crafted feature selection or the incorporation of guiding heuristics based on existing physics on oncologic principles. Alternatively, *post-hoc* methods are those that may be applied following model design and training. Improved MEI through these and other methods hold promise to promote the safety and quality of ML models in oncology as approaches grow in complexity, thereby improving synergy with clinicians and leading to real-world improvements in patient care. This collection includes five articles covering the following themes:

MR image reconstruction using patient-specific prior

MR imaging plays a critical role in many facets of oncology including diagnosis, treatment, and response assessment. When applied to advanced radiotherapy guidance through MRI-guided linear accelerators, acquisition time is of critical importance to reduce treatment time and mitigate effects of patient motion. Moreover, a high level of trust and comprehensive understanding of image features are imperative to facilitate critical therapeutic decisions. Grandinetti et al. propose a DL-based approach to enable high-quality reconstruction of under-sampled MR images enabling significant reduction in acquisition time. The authors propose incorporating patient-specific regularization using fully sampled pre-treatment diagnostic MRI of the same patient, providing the end user with an interpretable model overcoming limitations in population-average data. Using this approach, the authors demonstrate the ability to achieve high quality image reconstruction with significantly under-sampled data in both phantoms and patient cases.

Ultrasound-based prostate segmentation using an interpretable model expression

Trans-rectal ultrasound (TRUS) is commonly employed for the localization and guidance of needles for prostate biopsy and delivery of therapies including brachytherapy. In the case of ultrasound brachytherapy, localization and delineation of the needles and prostate gland are critical for effective treatment planning and delivery. In this application, time-efficiency is also critical since all steps must be completed while the patient is anesthetized. Peng et al. propose a semi-automatic prostate segmentation approach using ML and a principle curve based on an interpretable mathematical model expression. In this case, an ML model is combined with human initialization to create seed points, thereby providing flexibility for human input while augmenting results with an ML model providing the user with understanding and control of the algorithm output. The authors demonstrate improved prostate segmentation accuracy on patient images compared to existing state-of-the-art algorithms.

Cancer detection using non-invasive behavioral data

Several ML approaches have been proposed to aid in the early detection of cancer based on non-invasive data such as self-reported lifestyle characteristics, weight, or heart rate. However, ML models derived from large patient cohorts with potentially noisy or extraneous features may preclude interpretability by end users. Jiang et al. describe an ML model trained to predict gastric cancer diagnosis based on non-invasive lifestyle characteristics such as age, smoking history, and family cancer history. To ensure interpretability, the authors choose decision tree classifiers enabling *post hoc* analysis of input feature importance, demonstrating that XGBoost provided the highest prediction performance in the test cohort. Feature importance was reported, demonstrating that the top 5 features matched with those reported as predicted in previous prospective studies.

Outcome prediction using a human-in-the-loop-based Bayesian network approach

Outcome prediction for hepatocellular carcinoma (HCC) patients following stereotactic body radiotherapy (SBRT) remains challenging due to limitations in curating training databases and overcoming imbalances in outcomes, which can lead to biased prediction results. Luo et al. propose a "human-in-the-loop" (HITL) based Bayesian network approach to mitigate these challenges by including input from human experts during the selection of clinical features derived from HCC patients to create the prediction model of post-SBRT albumin-bilirubin grades, rather than relying on purely algorithmic feature selection which may suffer from biases in the training data. The HITL was found to not only improve interpretability of the final models, but also outperformed a purely data-driven approach in an independent test cohort from an outside institution.

Review of interpretable ML predictions for decision making in oncology

Following the application-specific investigations with considerations for MEI in the preceding four articles, Lu et al. conduct a thorough discussion and review of the challenges and importance of MEI in oncology, and an overview of algorithms and strategies applied to interpretation of complex non-linear ML models. The authors demonstrate the application of these

algorithms to identify cancerous breast masses using a publicly available dataset of cell nuclei characteristics, highlighting strengths and limitations of each approach. The authors conclude that MEI is an important consideration in oncology, and that many strategies exist to probe even highly complex and non-linear ML models, potentially leading to improvements in both performance and clinical adoption.

Author contributions

CW and TW are the coeditors for this Research Topic. WH is the review editor for this Research Topic. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine learning in oncology: A clinical appraisal. *Cancer Lett* (2020) 481:55–62. doi: 10.1016/j.canlet.2020.03.032

2. Wang G, Ye JC, Mueller K, Fessler JA. Image reconstruction is a new frontier of machine learning. *IEEE Trans Med Imag* (2018) 37(6):1289–96. doi: 10.1109/TMI.2018.2833635

3. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Comput Biol Med* (2020) 127:104065. doi: 10.1016/j.compbiomed.2020.104065

4. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Comput Struct Biotechnol J* (2018) 16:34–42. doi: 10.1016/j.csbj.2018.01.001

5. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med* (2020) 49(9):849–56. doi: 10.1111/jop.13042

6. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet* (2015) 16(6):321–32. doi: 10.1038/nrg3920

7. Elfiky AA, Pany MJ, Parikh RB, Obermeyer Z. Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw Open* (2018) 1(3):e180926. doi: 10.1001/jamanetworkopen.2018.0926

8. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine learning-based models for prediction of toxicity outcomes in radiotherapy. *Front Oncol* (2020) 10:790. doi: 10.3389/fonc.2020.00790

9. Chang M, Canseco JA, Nicholson KJ, Patel N, Vaccaro AR. The role of machine learning in spine surgery: The future is now. *Front Surg* (2020) 7:54. doi: 10.3389/fsurg.2020.00054

10. Jia X, Ren L, Cai J. Clinical implementation of AI technologies will require interpretable AI models. *Med Phys* (2020) 47(1):1–4. doi: 10.1002/mp.13891

11. Joshi G, Walambe R, Kotecha K. A review on explainability in multimodal deep neural nets. *IEEE Access.* (2021) 9:59800–21. doi: 10.1109/ACCESS.2021.3070212

12. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Sci (80-)* (2019) 363(6433):1287–9. doi: 10.1126/science.aaw4399

13. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PloS Med* (2018) 15(11):e1002683. doi: 10.1371/journal.pmed.1002683

14. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imag* (2020) 6(6):52. doi: 10.3390/jimaging6060052

15. Gulum MA, Trombley CM. Kantardzic m. a review of explainable deep learning cancer detection models in medical imaging. *Appl Sci* (2021) 11(10):4573. doi: 10.3390/app11104573

16. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. Available at: http://arxiv.org/abs/1312.6034.

17. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* (2017) 1705:07874.