Reply to A. Pfob et al

We thank Pfob et al[1] for their insightful perspective and recent contribution with respect to our article[2] in *JCO Clinical Cancer Informatics*. Indeed, they have raised important points about machine learning (ML) methods and modeling performances. Ultimately, robust data provenance, model design, and transparency in the reported frameworks will help demystify ML *black box models*.[3,4] In working toward standardizing practices, this will boost confidence in ML algorithms for clinical translation and applications. We would like to take this opportunity to reply to their thoughtful commentary.

Although ML models are not new, it is apparent that their applications across medical and scientific specialities are growing. Varying ML classifiers, feature sets, and model complexity can raise questions about selecting the optimal technique to handle prediction tasks. To streamline this process, there are recent initiatives to update standard reporting of ML frameworks; for example, the TRIPOD (Transparent Reporting of a Multivariable Prediction Model) statement is undergoing an update to include a TRIPOD-ML version.[5] This recent effort emphasizes the demand for increasing accessibility and understanding of ML approaches to clinicians. Additionally, data scientists are encouraged to report model parameters, including hyperparameters, kernels, and the preprocessing of data while constructing the model.[6,7]

Pfob et al draw attention to some key considerations for ML classification including (1) the importance of the input data, (2) influences of tuning hyperparameters, and (3) the need for external validation cohorts to yield generalizable inferences on new data sets. We agree with all those points. We further elaborate within the context of neoadjuvant therapy (NAT) for breast cancer below.

First, it is important to note that model inferences are dependent on the underlying data (ie, the training set). In essence, clinical data can carry inherent biases. For example, a study population with predominantly triple-negative breast cancer, which demonstrates high rates of pathologic complete response to anthracycline-taxane NAT, will inherently skew the model's performance metrics, regardless of tuning parameters introduced into the algorithm's pipeline. This consideration is also important when partitioning the data (ie, random splitting) into the training and test sets. Overall, there is a risk of variable classification performances from random partitioning when comparing models; therefore, it is recommended that the same training or test sets are used for all experiments.

Second, classification models demonstrate varying architectures, with differing assumptions and limitations on handling the input data. In tuning classifi adjusting hyperparameter settings will affect predict values. Other considerations include feature reduction operations to handle high-dimensional data, for example, using principal component analysis, univariate analysis, sequential feature selection, or correlation-based methods. Not surprisingly, the final feature sets that are loaded into a given model will yield alternate decision boundaries and predicted labels.

Finally, the development of a decision support tool is a multistep process. Among important elements related to clinical translation, the use of standard reporting methods (eg, patient characteristics, radiology, and pathology), internal testing, and ultimately external model validation transcends good practices in ML methods. These approaches are also fundamental in conventional regression models. These are, indeed, some of the fundamental aspects of good clinical research and are achieved by collaboration with and executing well-designed prospective trials. Many trials start with single-institution data sets to provide a conceptual framework before reaching external validation, which we agree is a critical step for evaluation and deriving robust generalizations from any given model. We hope to collaborate in the future to align interests in ML processes, centered on the long-term goal of guiding treatment strategies during NAT and in the adjuvant setting. These are common objectives in other domains, such as genomics, radiomics, and other *omics* approaches.

Notwithstanding these important considerations, we believe that ML frameworks have the potential to transform precision oncology, but we agree that this should be coupled with calculated optimism. The overall potential impact includes high-performing prediction and prognostic models that accurately represent real-world patient cases and outcomes, yield meaningful clinical signatures, and build upon the work of our colleagues to establish cogent methods. Overall, we share the same vision of improving patient care.

In conclusion, we thank Pfob et al for initiating some important discussions surrounding ML frameworks. Their expertise and insight have kindled deeper reflection on future ML work.

Nicholas Meti, MD
Division of Medical Oncology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada

Ali Sadeghi-Naini, PhD
Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, Ontario, Canada
Department of Radiation Oncology, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada

William T. Tran, MRT(T), MSc, PhD
Department of Radiation Oncology, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada
Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada

CORRESPONDING AUTHOR

William T. Tran, MRT(T), MSc, PhD, Department of Radiation Oncology, Sunnybrook Health Sciences Centre, 2075 Bayview Ave (Room TB-97), Toronto, Ontario, Canada M4N 3M5; e-mail: william.tran@sunnybrook.ca.

SUPPORT

W.T.T. received grant funding from the Tri-Council (CIHR) Government of Canada's New Frontiers in Research Fund (NFRF, Grant No. NFRFE-2019-00193), the Terry Fox Research Institute (TFRI, Grant No. 1083), and the Women's Health Golf Classic Foundation Fund. A.S.N.'s laboratory is funded by the TFRI (Grant No. 1083), NFRF (Grant No. NFRFE-2019-00193), and the Natural Sciences and Engineering Research Council of Canada (NSERC, Grant Nos RGPIN-2016-06472 and CRDPJ507521-16).

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by the authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Ali Sadeghi-Naini

Research Funding: Elekta

Patents, Royalties, Other Intellectual Property: I have two patents on application of quantitative ultrasound imaging for tissue characterization and therapy response monitoring

No other potential conflicts of interest were reported.

REFERENCES

1. Pfob A, Sidey-Gibbons C, Heil J: Response prediction to neoadjuvant systemic treatment in breast cancer—Yet another algorithm? JCO Clin Cancer Inform 5:654-655, 2021
2. Meti N, Saednia K, Lagree A, et al: Machine learning frameworks to predict neoadjuvant chemotherapy response in breast cancer using clinical and pathological features. JCO Clin Cancer Inform 5:66-80, 2021
3. Cabitza F, Rasoini R, Gensini GF: Unintended consequences of machine learning in medicine. JAMA 318:517, 2017
4. Wang F, Casalino LP, Khullar D: Deep learning in medicine—Promise, progress, and challenges. JAMA Intern Med 179:293, 2019
5. Collins GS, Moons KGM: Reporting of artificial intelligence prediction models. Lancet 393:1577-1579, 2019
6. Haibe-Kains B, Adam GA, Hosny A, et al: Transparency and reproducibility in artificial intelligence. Nature 586:E14-E16, 2020
7. Luo G: A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Netw Model Anal Health Inform Bioinform 5:18, 2016

DOI: https://doi.org/10.1200/CCI.21.00059; published at ascopubs.org/journal/cci on June 10, 2021.