Ann Surg Oncol
https://doi.org/10.1245/s10434-023-14144-5

Annals of SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ASO PERSPECTIVES

Rethinking Risk Modeling with Machine Learning

Adam Yala, PhD[1,2], and Kevin S. Hughes, MD[3]

[1]UC Berkeley, Berkeley; [2]UCSF, San Francisco; [3]Surgical Oncology, Medical University of South Carolina, Charleston

Accurate risk assessment is essential for the early detection and prevention of breast cancer. With the foresight offered by risk models, high-risk patients can benefit from supplemental imaging, more frequent screening, and chemoprevention to improve their outcomes. Similarly, low-risk patients can be guided toward longer screening intervals and avoid overtreatment. As such, there have been considerable investments in the development of risk-based guidelines for supplemental imaging, personalized screening frequency, and chemoprevention.[1–4] However, the risk models underlying these national efforts give gross, generalized risk estimates that are inaccurate at the individual level, limiting the efficacy of existing guidelines. For instance, current National Comprehensive Cancer Network (NCCN) guidelines recommend supplemental magnetic resonance imaging (MRI) for patients with 20% or greater lifetime risk of breast cancer.[5] However, under these guidelines, more than 97% of supplemental screening MRIs will not detect cancer,[6] indicating that most of these patients did not need MRIs. Conversely, only 25% of patients with breast cancer will be eligible for MRI before their diagnosis, indicating a missed opportunity for 75% of patients with cancer.[7] Guidelines for chemoprevention and screening frequency are similarly inefficient. These challenges stem from the limitations of the guideline's underlying risk models. Improving predictors of individual cancer risk remains essential to improving the systematic early detection and prevention of breast cancer.

In a recent *Journal of Clinical Oncology* article, Eriksson et al.[8] have demonstrated that image-derived risk models, which were previously shown to outperform the Tyrer–Cuzick model in short-term risk estimation (2-year risk),[9] also outperform the baseline in long-term risk estimation (10-year risk). Specifically, the study demonstrated that an image-based risk model, based on prespecified mammographic features, outperformed the Tyrer–Cuzick v8 model on a case-control cohort across a 10-year period. Throughout their 10-year follow-up window, they found that 20% of all women with breast cancer were deemed as high risk by their image model, compared with 7.1% by Tyrer–Cuzick v8. While this result may be confounded by good performance cancers earlier in the observation window, the study remains promising. This improved capacity to predict long-term risk is especially important to support the primary prevention of breast cancer, as tumor development is estimated to take 5–20 years. By extending the successes of image-based risk modeling[9–14] to long-term risk estimation, Eriksson et al. contribute to a larger paradigm shift in risk modeling.

The traditional approach for developing risk predictors, as exemplified by the Tyrer–Cuzick model,[15] relies on expert knowledge to identify key risk factors. These curated risk factors, (e.g., patient age, family history, mammographic density, etc.), are then combined in statistical models to estimate breast cancer risk. While traditional tools such as the Tyrer–Cuzick model are widely adopted, the models demonstrate limited performance at the individual level. Moreover, it has proven difficult for experts to improve these tools with new risk factors, suggesting that this approach may have already reached its limits. For instance, investigators have extensively explored mammographic breast density as a marker of risk; however, the performance of the Gail and Tyrer–Cuzick models improved only marginally after the introduction of this factor, obtaining areas under the curve (AUCs) of 0.61 and 0.59 compared with 0.57 and 0.55, respectively.[16] The manual identification of risk factors remains a critical bottleneck in improvement of risk models.

Recently, radiomics approaches,[17–19] which evaluate combinations of texture or shape features, have emerged as a popular direction in medical imaging. Radiomics models promise to improve the flexibility of traditional risk models

First Received: 15 June 2023
Accepted: 31 July 2023

A. Yala, PhD
e-mail: yala@berkeley.edu

Published online: 13 August 2023

by adding expert-defined, yet automatically measured, imaging features. While significantly improved, this approach is still fundamentally limited by features selected by the investigator, restricting the ability of the developed tools to discover new features directly from the patient data.

Artificial intelligence (AI) methods, which can operate directly on full resolution images and leverage any predictive pattern available, offer a paradigm shift in the development of risk models. Instead of relying on investigator ingenuity, AI approaches allow us to view risk modeling as an optimization problem. In doing so, AI tools can uncover risk cues unknown to human readers directly from the data, allowing them to reap the full potential of the modality.

This AI approach has already transformed short-term risk prediction, with models such as Mirai[7] and Sybil[10] achieving state-of-the-art performance in 5-year breast cancer and 6-year lung cancer risk, respectively, in large international external validation studies.[10,11] In this study, Eriksson et al. further this general paradigm shift, extending these successes to long-term risk estimation. Like prior studies in AI-derived risk assessment, the model in this paper did not outperform the Tyrer–Cuzick model because it had access to more data; as both methods used mammograms in some fashion. Instead, the methods differ in how they leveraged the mammograms. By detecting and combining subtle features of the mammogram, the image-based model was able to significantly outperform the Tyrer–Cuzick model, which only benefits from breast density. These results reinforce the exciting promise of AI methods to transform risk assessment. Moreover, while progress in traditional risk models has stagnated, AI-driven methods have the potential for further dramatic improvement. Current AI methods only leverage a single episode of mammography to assess cancer risk, which only touches a fraction of the rich multi-modal and longitudinal imaging available. Ever larger multimodal datasets, improved algorithms, and increased computing power all have the promise to further advance AI risk assessment.

To realize the promise of AI-driven risk models, future work requires strengthened standards of rigor to ensure methodological progress. Specifically, studies should compare their proposed methods to published prior work, establishing means to gauge technical improvements. While Eriksson et al. focused on a single commercial model for their study, other image-based risk models, including Mirai, are publicly available. We believe the study would have benefited from wider benchmarking. Similarly, the study would have benefited from validating their results on more diverse datasets, such as EMBED.[20] Currently, the study only evaluates their risk model on the same screening cohort used to develop the model; as a result, it remains difficult to gauge the external validity of the current results to more diverse populations. Finally, AI model developers should make their tools easily available to other researchers to enable new work to compare against their approaches. Open benchmarking and globally diverse validation efforts are critical to ensuring that AI methods for cancer risk assessment are actually improving.

More clinical research is also needed to translate the advancements in cancer risk modeling into tangible advances in care. In both this study and prior work validating Mirai, AI risk scores show higher accuracies in the near term (within 2 years) than far term (5–10 years), suggesting the models are partially capturing cancers already present within the mammogram but not detected by the radiologist. These trends, in addition to the overall improved accuracy, suggest that leveraging these models to decide supplemental imaging, instead of current lifetime risk measures,[5] would be more effective in reducing interval cancers. Moreover, while lifetime risk measures disproportionately exclude older women, image-based risk tools are not inherently skewed by patient age[21] and they can identify the near-term risk signal most relevant for personalized screening. Given that AI risk scores remain more accurate than traditional measures after 10 years,[8] chemoprevention guidelines could also benefit from AI risk tools. While the broad opportunities for AI tools to improve guidelines for cancer prevention and screening are clear, the optimal clinical protocols for each of these use cases remain unclear. Prospective studies with diverse populations are needed to advance clinical guidelines and to achieve the broad promise[22–24] of AI in medicine. The future is AI-derived risk estimation and AI-powered clinical guidelines, and the faster we get there, the better our patients will be cared for.

DISCLOSURE Kevin Hughes Honoraria: Receives Honoraria from TME (Targeted Medical Education, genetics education and consulting), MedNeon-Invitae (Genetic testing/education subdivision) 23&Me(Genetics company), Invitae(Genetics company), Ambry (Genetic testing) and Astra Zeneca (Parmaceutical Company) Financial Interest: Has a founded and has a financial interest in CRA Health (Formerly Hughes RiskApps) which was acquired by Volpara in January, 2021. CRA Health develops risk assessment models/software with a particular focus on breast cancer and colorectal cancer. Dr. Hughes is the Co-Creator of Ask2Me.Org which is freely available for clinical use and is licensed for commercial use by the Dana Farber Cancer Institute and the Massachusetts General Hospital (MGH). Dr. Hughes's interests in CRA Health and Ask2Me.Org were reviewed and are managed by Massachusetts General Hospital and Mass General-Brigham in accordance with their conflict of interest policies.

REFERENCES

1. Esserman LJ; WISDOM Study and Athena Investigators. The WISDOM study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer*. 2017;13(3):34. https://doi.org/10.1038/s41523-017-0035-5.

2. Lee C, McCaskill-Stevens W. Tomosynthesis mammographic Imaging Screening Trial (TMIST): an invitation and opportunity for the national medical association community to shape the future of precision screening for breast cancer. *J Natl Med Assoc*. 2020;112(6):613–8. https://doi.org/10.1016/j.jnma.2020.05.021.
3. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, Sroczynski G, Hall P, Cuzick J, Evans DG, Simard J, Garcia-Closas M, Schmutzler R, Wegwarth O, Pharoah P, Moorthie S, De Montgolfier S, Baron C, Herceg Z, Turnbull C, Balleyguier C, Rossi PG, Wesseling J, Ritchie D, Tischkowitz M, Broeders M, Reisel D, Metspalu A, Callender T, de Koning H, Devilee P, Delaloge S, Schmidt MK, Widschwendter M. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol*. 2020;17(11):687–705. https://doi.org/10.1038/s41571-020-0388-9.
4. Roux A, Cholerton R, Sicsic J, Moumjid N, French DP, Giorgi Rossi P, Balleyguier C, Guindy M, Gilbert FJ, Burrion JB, Castells X, Ritchie D, Keatley D, Baron C, Delaloge S, de Montgolfier S. Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the "My Personal Breast Screening" (MyPeBS) randomised clinical trial. *BMC Cancer*. 2022;22(1):507. https://doi.org/10.1186/s12885-022-09484-6. PMID:35524202;PMCID:PMC9073478.
5. National Comprehensive Cancer Network: Breast Cancer Screening. (Jan 2022 version) https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf
6. Vreemann S, Gubern-Mérida A, Schlooz-Vries MS, Bult P, van Gils CH, Hoogerbrugge N, Karssemeijer N, Mann RM. Influence of risk category and screening round on the performance of an MR imaging and mammography screening program in carriers of the brca mutation and other women at increased risk. *Radiology*. 2018;286(2):443–51. https://doi.org/10.1148/radiol.2017170458.
7. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan YL, Lamb L, Hughes K, Lehman C, Barzilay R. Toward robust mammography-based models for breast cancer risk. *Sci Transl Med*. 2021;13(578):eaba4373. https://doi.org/10.1126/scitranslmed.aba4373.
8. Eriksson M, Czene K, Vachon C, Conant EF, Hall P. Long-term performance of an image-based short-term risk model for breast cancer. *J Clin Oncol*. 2023;41(14):2536–45.
9. Eriksson M, Czene K, Strand F, et al. Identification of women at high risk of breast cancer who need supplemental screening. *Radiology*. 2020;297:327–33.
10. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, Bourgouin PP, Chan P, Mrah S, Amayri W, Juan YH, Yang CT, Wan YL, Lin G, Sequist LV, Fintelmann FJ, Sybil Barzilay R. A validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. *J Clin Oncol*. 2023;41(12):2191–200. https://doi.org/10.1200/JCO.22.01345.
11. Yala A, Mikhael PG, Strand F, Lin G, Satuluru S, Kim T, Banerjee I, Gichoya J, Trivedi H, Lehman CD, Hughes K, Sheedy DJ, Matthis LM, Karunakaran B, Hegarty KE, Sabino S, Silva TB, Evangelista MC, Caron RF, Souza B, Mauad EC, Patalon T, Handelman-Gotlib S, Guindy M, Barzilay R. Multi-institutional validation of a mammography-based breast cancer risk model. *J Clin Oncol*. 2022;40(16):1732–40. https://doi.org/10.1200/JCO.21.01337.
12. Dembrower K, Liu Y, Azizpour H, Eklund M, Smith K, Lindholm P, Strand F. Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. *Radiology*. 2020;294(2):265–72.
13. Zhu X, Wolfgruber TK, Leong L, Jensen M, Scott C, Winham S, Shepherd JA. Deep learning predicts interval and screening-detected cancer from screening mammograms: a case-case-control study in 6369 women. *Radiology*. 2021;301(3):550–8.
14. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*. 2019;292(1):60–6.
15. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111–30.
16. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, Fox L, Sergeant JC, Harvie MN, Wilson M, Beetles U, Gadde S, Lim Y, Jain A, Bundred S, Barr N, Reece V, Howell A, Cuzick J, Evans DG. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. 2015;17(1):147.
17. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction. *Semin Cancer Biol*. 2021;72:238–50. https://doi.org/10.1016/j.semcancer.2020.04.002.
18. Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, Gu Y, Li A, Lu N, He Z, Yang Y, Chen K, Ma J, Li C, Ma M, Li X, Zhang R, Zhong H, Ou Q, Zhang Y, He Y, Li G, Wu Z, Su F, Song E, Yao H. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. *JAMA Netw Open*. 2020;3(12):e2028086. https://doi.org/10.1001/jamanetworkopen.2020.28086.
19. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Staudt LM. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(25):1937–47.
20. Jeong JJ, Vey BL, Bhimireddy A, Kim T, Santos T, Correa R, Dutt R, Mosunjac M, Oprea-Ilies G, Smith G, Woo M, McAdams CR, Newell MS, Banerjee I, Gichoya J, Trivedi H. The EMory Br East imaging Dataset (EMBED): a racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. *Radiol Artif Intell*. 2023;5(1):e220047. https://doi.org/10.1148/ryai.220047.
21. Coopey SB, Acar A, Griffin M, Cintolo-Gonzalez J, Semine A, Hughes KS. The impact of patient age on breast cancer risk prediction models. *Breast J*. 2018;24(4):592–8.
22. Hassan AM, Nelson JA, Coert JH, Mehrara BJ, Selber JC. Exploring the potential of artificial intelligence in surgery: insights from a conversation with ChatGPT. *Ann Surg Oncol*. 2023;30(7):3875–8. https://doi.org/10.1245/s10434-023-13347-0.
23. Topol Eric J. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
24. Rajpurkar Pranav, Chen Emma, Banerjee Oishi, Topol Eric J. AI in health and medicine. *Nat Med*. 2022;28(1):31–8.