Harnessing the Power of Deep Learning to Assess Breast Cancer Risk

Manisha Bahl, MD, MPH

Dr Manisha Bahl is a radiologist at the Massachusetts General Hospital and an assistant professor of radiology at Harvard Medical School. Her research interests focus on applications of artificial intelligence to breast imaging and the assessment of digital breast tomosynthesis in the clinical setting. Dr Bahl is a prior recipient of an RSNA Research Scholar Grant and is currently the principal investigator of an NIH grant.

Accurate assessment of a woman's risk for breast cancer is necessary to guide informed decision making regarding personalized screening and prevention strategies. Most risk prediction models provide risk estimates at the population level but are not as precise at the individual level (1). Existing risk prediction models, such as the Gail and Tyrer-Cuzick models, are largely based on information gleaned from questionnaires, such as age at menarche, hormone replacement therapy use, and family history of breast cancer (1). Image-based information has not been incorporated into risk prediction models until relatively recently, when mammographic breast density (which refers to the amount of fibroglandular tissue in the breast) began receiving much attention as a risk factor (2). However, the use of breast density alone to estimate cancer risk is limited, as there can be interreader variability (if assessed by a radiologist rather than automated software), and it is unlikely to capture all of the rich information contained within a mammographic image (3).

In this issue of *Radiology*, Dembrower and colleagues (4) harnessed the power of deep learning (DL) with convolutional neural networks to determine individual breast cancer risk scores based on mammographic images. The recent growth of DL has been spurred by the increased availability of large data sets and advanced computer algorithms, in addition to increased computing power (5,6). While the traditional machine learning approach for classifying images is based on handcrafted features, DL uses high-level imaging features from large data sets and is based on networks of interconnected units (5). These units connect to form multiple layers (hence "deep") between the input and output layers, some of which are hidden, that can generate increasingly high-level representations of the inputted data (which are images) (7). The term *neural*

networks is inspired by the connectivity of neurons in the brain, and a convolutional neural network is a specific type of neural network used for image analysis (8).

In the study by Dembrower et al, a convolutional neural network was trained to estimate breast cancer risk by using cancer-free mammographic images from nearly 10 000 women in the Stockholm county area, 891 of whom were subsequently diagnosed with breast cancer (4). The DL model was validated with mammographic images from almost 1800 women. The authors present the results of the DL model tested on 2283 women (who were not represented in the training and validation sets), 278 of whom were subsequently diagnosed with breast cancer (4). The output of the model was the DL risk score, which reflects the likelihood of developing breast cancer based on a single mammographic image. The DL risk score ranges from 0 to 1, with 0 corresponding to no risk of developing cancer and 1 corresponding to maximum risk. The mean DL risk score from all four of a patient's mammographic views (ie, bilateral craniocaudal and mediolateral oblique views) was used to estimate an individual's risk level. The degree of correlation between the DL risk score and two density-based measurements (made by automated software) was found to be low to moderate (with Spearman correlation coefficients between 0.25 and 0.42), suggesting that the DL risk score is not simply a proxy for breast density.

Dembrower et al (4) compared the accuracy of the image-based DL model to that of two different models based on breast density (with density measurements made by automated software). In the test set of 2283 women, the DL model demonstrated a higher age-adjusted risk association for breast cancer compared with the best mammographic density model (odds ratios of 1.6 and 1.3, respectively; $P < .001$). Furthermore, the area under the receiver operating characteristic curve (AUC) for the DL model was higher than that for a model based on patient age and dense area (an automated measurement of breast density) (0.65 vs 0.60, respectively; $P < .001$). These findings suggest that a DL model trained with mammographic images and breast cancer outcome could more accurately predict breast cancer risk than density-based models.

To further evaluate the DL model's performance, Dembrower et al examined false-negative predictions made by the model (4). The authors defined a false-negative prediction as a DL risk score below the median when, in fact,

From the Department of Radiology, Massachusetts General Hospital, 55 Fruit St, WAC 240, Boston, MA 02114. Received November 5, 2019; revision requested November 14; revision received November 15; accepted November 18. Address correspondence to the author (e-mail: mbahl1@mgh.harvard.edu).

Conflicts of interest are listed at the end of this article.

See also the article by Dembrower et al in this issue.

the patient is subsequently diagnosed with breast cancer. The median is an arbitrary cut-off for this particular analysis but could vary depending on the specific setting in which the DL model is being used. The DL model was found to have a lower false-negative rate than the best mammographic density model (31% vs 36%, respectively; $P < .01$). This difference was most pronounced for women subsequently diagnosed with more aggressive cancers (eg, lymph node–positive cancers). This finding may be of clinical importance because lymph node status is an important prognostic factor, and women at risk for aggressive cancers may benefit from personalized and intensive screening regimens.

The study suggests that mammographic images contain indicators of risk not captured with use of breast density alone and perhaps beyond the limits of human detection. Herein lies the power of DL: to discover useful features that may not be discernible by the most experienced and skillful breast imagers and/or that are not currently known (8). The study by Dembrower et al is in line with other recent studies on the use of DL models as predictors of breast cancer risk (9,10). For example, Ha et al (9) found that an image-based DL model had greater predictive potential than breast density (odds ratios of 4.4 vs 1.7, respectively). The DL model achieved an overall accuracy of 72% among women who were subsequently diagnosed with breast cancer (9). Yala et al (10) found that an image-based DL model outperformed the Tyrer-Cuzick model, which is used in clinical practice (AUC of 0.68 vs 0.62, respectively; $P < .01$). The best performance was achieved by a model that incorporated both mammographic images and traditional risk factors (AUC of 0.70) (10). These studies demonstrate that image-based DL models offer promise as more accurate predictors of breast cancer risk than density-based models and existing epidemiology-based models.

Image-based DL models are based on the rich information contained within a mammographic image, and not on the subjectivity and variability inherent in human-made assessments of imaging features such as breast density. Future image-based DL models may be further strengthened by incorporating other sources of information, such as genomic data, although higher AUCs (ie, ≥ 0.8) may not be achievable for breast cancer risk prediction models. Further work could also provide insight into the types of imaging patterns that DL models are using to

predict breast cancer risk. Yala et al (10) hypothesize that image-based DL models may rely on the orientations of different fine-grain tissue patterns relative to global patterns in a patient's breast. Heat maps or saliency maps could also be used to identify regions within the mammographic images most commonly encountered in patients with low (eg, blue) and high (eg, red) risk of breast cancer (9). It remains uncertain, however, whether image-based DL methods will ever be fully understandable by us humans. Although such models could potentially replace existing risk prediction models, further research is needed to thoroughly validate them across mammography vendors and institutions. Continuing work is therefore needed to strengthen and evaluate risk models to support personalized screening and prevention strategies and ultimately reduce the burden of breast cancer.

Disclosures of Conflicts of Interest: disclosed no relevant relationships.

References

1. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst* 2010;102(10):680–691.
2. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. *JAMA Oncol* 2018;4(9):e180174.
3. Sprague BL, Conant EF, Onega T, et al. Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. *Ann Intern Med* 2016;165(7):457–464.
4. Dembrower K, Liu Y, Azizpour H, et al. Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. *Radiology* 2020;294:265–272.
5. Burt JR, Torosdagli N, Khosravan N, et al. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. *Br J Radiol* 2018;91(1089):20170545.
6. Abdelhafiz D, Yang C, Ammar R, Nabavi S. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinformatics* 2019;20(Suppl 11):281.
7. Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging* 2019;49(4):939–954.
8. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017;208(4):754–760.
9. Ha R, Chang P, Karcich J, et al. Convolutional neural network based breast cancer risk stratification using a mammographic dataset. *Acad Radiol* 2019;26(4):544–549.
10. 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):60–66.