ORIGINAL RESEARCH • BREAST IMAGING

Radiology

Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation

Constance D. Lehman, MD, PhD • Adam Yala, MEng • Tal Schuster, MSc • Brian Dontchos, MD • Manisha Bahl, MD, MPH • Kyle Swanson, BS • Regina Barzilay, PhD

From the Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Avon Comprehensive Breast Evaluation Center, 55 Fruit St, WAC 240, Boston, MA 02114-2698 (C.D.L., B.D., M.B.); and Massachusetts Institute of Technology, Cambridge, Mass (A.Y., T.S., K.S., R.B.). Received March 24, 2018; revision requested May 14; revision received August 21; accepted August 27. **Address correspondence to** C.D.L. (e-mail: *clehman@partners.org*).

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

See also the editorial by Chan and Helvie in this issue.

Radiology 2019; 290:52–58 • https://doi.org/10.1148/radiol.2018180694 • Content code: BR

Purpose: To develop a deep learning (DL) algorithm to assess mammographic breast density.

Materials and Methods: In this retrospective study, a deep convolutional neural network was trained to assess Breast Imaging Reporting and Data System (BI-RADS) breast density based on the original interpretation by an experienced radiologist of 41 479 digital screening mammograms obtained in 27 684 women from January 2009 to May 2011. The resulting algorithm was tested on a held-out test set of 8677 mammograms in 5741 women. In addition, five radiologists performed a reader study on 500 mammograms randomly selected from the test set. Finally, the algorithm was implemented in routine clinical practice, where eight radiologists reviewed 10 763 consecutive mammograms assessed with the model. Agreement on BI-RADS category for the DL model and for three sets of readings—*(a)* radiologists in the test set, *(b)* radiologists working in consensus in the reader study set, and *(c)* radiologists in the clinical implementation set—were estimated with linear-weighted κ statistics and were compared across 5000 bootstrap samples to assess significance.

Results: The DL model showed good agreement with radiologists in the test set (κ = 0.67; 95% confidence interval [CI]: 0.66, 0.68) and with radiologists in consensus in the reader study set (κ = 0.78; 95% CI: 0.73, 0.82). There was very good agreement (κ = 0.85; 95% CI: 0.84, 0.86) with radiologists in the clinical implementation set; for binary categorization of dense or nondense breasts, 10 149 of 10 763 (94%; 95% CI: 94%, 95%) DL assessments were accepted by the interpreting radiologist.

Conclusion: This DL model can be used to assess mammographic breast density at the level of an experienced mammographer.

Online supplemental material is available for this article.

Mammographic breast density can mask cancers at mammography and is an independent risk factor for breast cancer (1–3). Legislation mandating patients be notified of mammographic breast density has passed in more than 30 states, and a federal bill is under consideration. Details of state legislation vary, but most states require direct reporting to the patient that breast density can mask cancers at mammography and that the patient may benefit from additional testing.

Qualitative assessment of mammographic breast density is subjective and varies widely between radiologists (4–10). In a study of 83 radiologists who assessed breast density, Sprague et al (4) found extreme variation in qualitative density assessment per the Breast Imaging Reporting and Data System (BI-RADS), with 6%–85% of mammograms assessed as either heterogeneously or extremely dense depending on radiologist interpretation. In a study of 34 radiologists, the intraradiologist agreement of density assessments among women who underwent two examinations varied from 62% to 87% (6).

Commercially available methods for automated assessment of breast density do exist; however, they yield mixed results in agreement with expert qualitative density assessments, with κ scores of 0.32–0.61 (11,12). These methods tend to result in over- or underreporting of breast density when compared with qualitative assessment by radiologists (11,13). A recent study found significant differences in density assessments in the same 4170 women with two software programs (Volpara, Volpara Solutions, Wellington, New Zealand; Quantra, Hologic, Bedford, Mass), with the software programs showing 37% and 51%, respectively, of women had dense breast tissue. In the same set of mammograms, radiologists determined 43% of the women had dense breast tissue (13).

Deep learning (DL) has been gaining traction in radiology (12,14–17). Specifically, there has been preliminary work with DL methods to assess breast density (12,18); however, none of these techniques have been implemented in clinical practice, raising questions about clinical acceptance by practicing radiologists and the effect on patient care. In contrast, our purpose was to develop a DL algorithm we could use to reliably assess breast density and to measure the acceptance of its predictions in real-time clinical practice. We hypothesize that DL models can be applied to assess breast density at the same level as experienced breast imagers and that they can be accepted into routine clinical practice.

Abbreviations

BI-RADS = Breast Imaging Reporting and Data System, CI = confidence interval, DL = deep learning

Summary

A deep learning algorithm was used to assess mammographic breast density at the level of an experienced mammographer during routine clinical practice.

Implications for Patient Care

- A deep learning algorithm was used to reliably and accurately assess mammographic breast density in a large clinical practice.
- Given the high level of agreement between the deep learning algorithm and experienced mammographers, this algorithm has the potential to standardize and automate routine breast density assessment.

Materials and Methods

This retrospective study was approved by the Massachusetts General Hospital institutional review board, who waived the need to obtain informed consent, and was compliant with the Health Insurance Portability and Accountability Act.

Development and Testing of the DL Model

We developed and tested our DL model by using 58 894 randomly selected digital mammograms from 39 272 women screened between January 2009 and May 2011; there were no exclusion criteria (eg, prior surgery, implants, etc). The women were randomly assigned to a training (n = 41 479), development (n = 8738), or test (n = 8677) set. Breast density was recorded by one of 12 radiologists who specialized in breast imaging and who had 5–33 years of experience following the American College of Radiology BI-RADS lexicon (category a, almost entirely fatty; category b, scattered areas of fibroglandular tissue; category c, heterogeneously dense; category d, extremely dense) (19).

We implemented our model by using a deep convolutional neural network, ResNet-18 (20), with PyTorch (2018, version 0.31; *pytorch.org*). The model was trained to map images in a single view, without any exclusions, to assess breast density. To aggregate the density assessments from each view into an assessment for the examination, we used the consensus density across views, if present; otherwise, ties were broken randomly. During model development, we augmented our training data with random flips and rotations of the original images and experimented with various regularization strategies and model architectures. We chose this network because it had the best performance in the development set. Appendix E1 (online) contains additional details regarding our DL model.

After training, we assessed the proportion of examinations in which the model enabled prediction of the density rating given by the original interpreting radiologist for the held-out test set. We evaluated binary categorization for BI-RADS categories c and d (dense) or a and b (nondense) and for the four individual BI-RADS categories (categories a, b, c, d). For both binary and four-way categorizations, we quantified the types of disagreements made by the model in a confusion matrix. Each examination is placed in a specific cell, as determined by the density assessment of the radiologist (row) and model (column). Percentages were computed by dividing the number of examinations in the cell by the number of examinations in the row. We also measured agreement between our DL model and the original radiologist assessment across the four BI-RADS categories.

Reader Study

We recorded breast density assessments by five breast imagers with 2–23 years of experience (C.D.L., B.D., M.B., with 23, 5, and 3 years of experience, respectively) for 500 mammograms randomly selected from the test set. The breast imagers were blinded to each other, to the original radiologist's interpretation, and to the DL model assessment. We compared agreement between the DL model and the consensus (majority assessment) of the five breast imagers and between the DL model and the original interpreting radiologist. We also compared agreement between the five breast imagers in consensus and the original interpreting radiologist. Because the DL model was trained on the assessments of multiple radiologists, we hypothesized that the DL model would show higher agreement with the assessment made by the radiologists working in consensus than with the original radiologist assessment.

Clinical Implementation and Acceptance of DL Assessment

We assessed the clinical acceptance of our DL model by using 10 763 consecutive screening digital mammograms from January to May of 2018. No mammographic examinations were excluded (eg, no exclusions due to prior surgery, implants, etc). Mammograms were automatically retrieved from the picture archiving and communication system and were processed with the DL algorithm; DL density assessment (BI-RADS category a, b, c, or d) was sent to a commercially available mammography reporting software program (Magview 2018, version 8.0.143; Magview. Burtonsville, Md). The time for retrieval and assessment of the mammogram and presentation of the density in the patient's mammography report took less than five seconds per case and was implemented for mammograms obtained from one local and five remote screening centers. Mammograms were analyzed at our mammography review workstations (SecurView Workstation; Hologic) by following our routine clinical workflow for assessment and reporting. Eight breast imagers with 2–23 years of experience participated in clinical implementation evaluation. None of these radiologists contributed density assessments to our training, development, or test sets. During the review, radiologists (C.D.L., B.D., M.B.) were provided with the DL model density assessment in the electronic report provided by the reporting software. The final density assessment of the mammogram was at the discretion of the radiologist (ie, to agree or disagree with the DL algorithm). We measured the proportion of automatic DL assessments accepted by the interpreting radiologist for the final reading, both for binary categorization as dense or nondense and across the four BI-RADS categories.

Statistical Analysis

The proportion of mammograms in which DL model assessment matched radiologist assessment in the test set and in the

Table 1: Patient Characteristics and Breast Density as Assessed by Original Interpreting Radiologists in Training and Test Sets

Characteristic	Training Set	Test Set
Screening mammograms	41 479	8677
No. of patients	27 684	5741
Age (y)*	57.4 (31–97)	57.5 (28–92)
<40	459 (1)	100 (1)
40–49	11 602 (28)	2398 (28)
50–59	12 235 (29)	2591 (30)
60–69	10 556 (25)	2171 (25)
≥70	6627 (16)	1417 (16)
Radiologist-assessed breast density		
Almost entirely fatty	3818 (9)	792 (9)
Scattered areas of fibroglandular tissue	20 913 (50)	4386 (51)
Heterogeneously dense	14 856 (36)	3096 (36)
Extremely dense	1892 (5)	403 (5)

Note.—Unless otherwise indicated, data are numbers of mammograms, and data in parentheses are percentages.
* Data are mean age, and data in parentheses are the range.

Table 2: Proportion of DL Model Assessments Matching Radiologist Assessments in the Test and Clinical Implementation Settings

Setting	Dense or Nondense	Four BI-RADS Categories	Total No. of Patients
Test set accuracy	7566 (87) [86, 88]	6681 (77) [76, 78]	8677
Clinical implementation set acceptance	10 149 (94) [94, 95]	9729 (90) [90, 91]	10 763

Note.—Data are number of patients. Data in parentheses are accuracy or acceptance, as indicated, and are percentages. Data in brackets are 95% confidence intervals. DL = deep learning.

clinical implementation set was estimated by using 95% Wilson confidence intervals. We quantified the types of disagreements in confusion matrices and computed agreement between final assessment and DL assessment across the four BI-RADS categories, estimating with weighted κ using linear weighting. The κ statistics were compared across 5000 bootstrap samples to assess significance. We computed all statistics using statistical software (scikit-learn, version 0.19.1; *scikit-learn.org*).

Results

The training set consisted of 27 684 women who underwent 41 479 screening mammographic examinations, with an average patient age of 57 years (age range, 31–97 years). The original interpreting radiologists assessed breast density as almost entirely fatty in 9% of examinations, as scattered areas of fibroglandular tissue in 50% of examinations, as heterogeneously dense in 36% of examinations, and as dense in 5% of examinations. The held-out test set consisted of 5741 women who underwent 8677 examinations, with an average patient age of 57.5 years (range, 28–92 years). The original interpreting radiologists assessed breast densities as almost entirely fatty in 9% of patients, as scattered areas of fibroglandular tissue in 50% of patients, as heterogeneously dense in 36% of patients, and as dense in 5% of patients (Table 1).

Testing of DL Model

We first describe the distribution of the DL algorithm assessments. Of the 8677 held-out test mammograms, the DL model assessed 8% as almost entirely fatty, 52% as scattered areas of fibroglandular tissue, 38% as heterogeneously dense, and 2% as extremely dense. For binary categorization of dense or nondense tissue, the density assigned by the DL model matched that assigned by the original interpreting radiologist for 7566 of 8677 (87%; 95% CI: 86%, 88%) mammograms. The model downgraded 567 of 8677 (7%) mammograms from dense to nondense and upgraded 547 of 8677 (6%) mammograms from nondense to dense. For four-way BI-RADS categorization, the DL model matched the radiologist interpretation for 6681 of 8677 (77%; 95% CI: 76%, 78%) mammograms (Table 2). Of the 1993 instances of disagreement between the DL model and human observers, 1105 (55%) were between scattered areas of fibroglandular tissue and heterogeneously dense, 566 (28%) were between almost entirely fatty and scattered areas of fibroglandular tissue, and 323 (16%) were between heterogeneously dense and extremely dense. Model disagreements between almost entirely fatty and extremely dense did not occur. (Fig 1). Agreement between density assessments with our DL model and those of the original interpreting radiologist was good (κ = 0.67; 95% CI: 0.66, 0.68) (Table 3).

Reader Study

We next describe the distribution density assessments of the readers in consensus, the original interpreting radiologist, and the DL model for the 500 random mammograms from the test set. The readers working in consensus assessed 13% of mam-

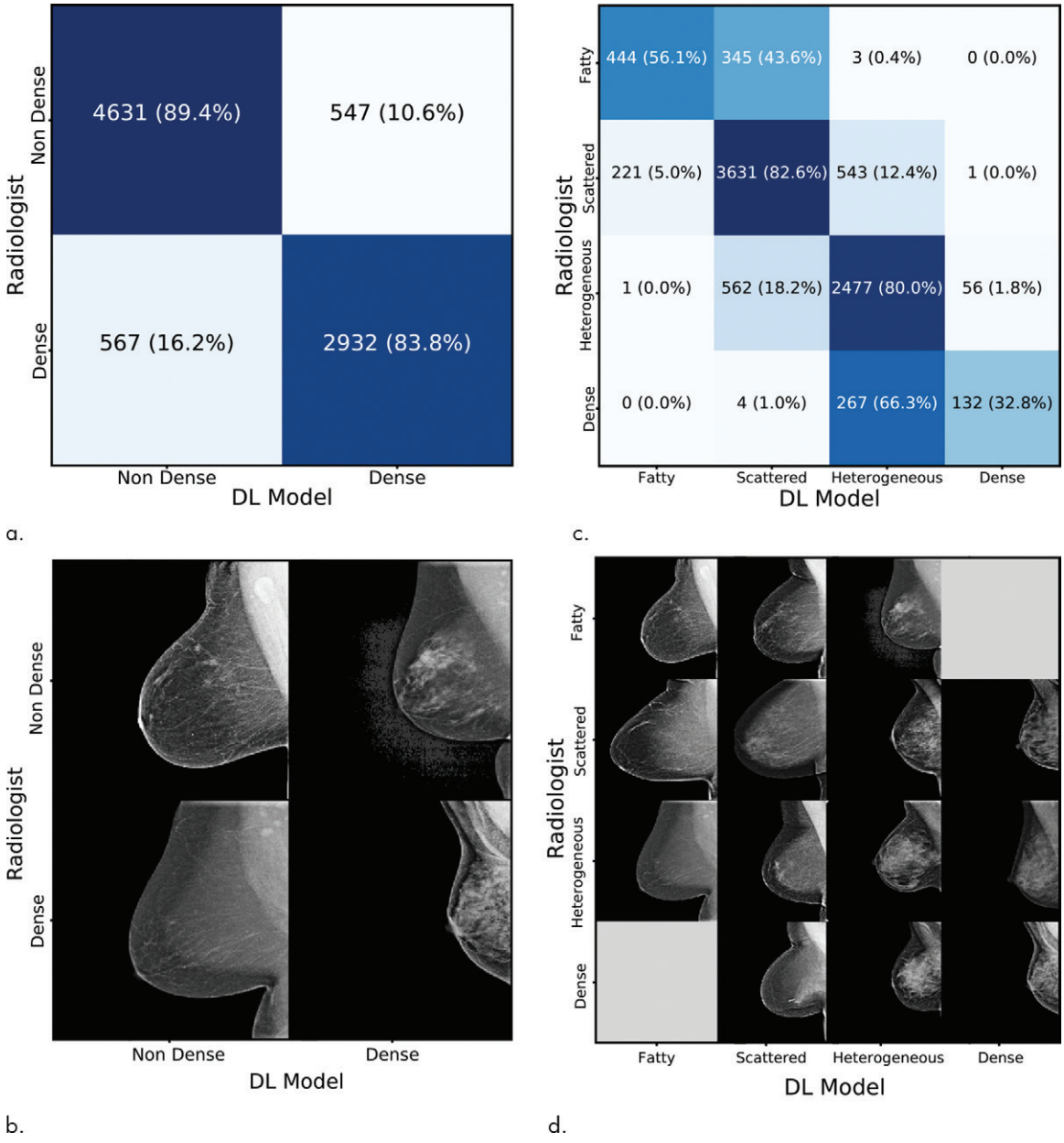

Figure 1: Test set assessment. Comparison of the original interpreting radiologist assessment with the deep learning (DL) model assessment for **(a)** binary and **(c)** four-way mammographic breast density classification. **(b, d)** Corresponding examples of mammograms with concordant and discordant assessments by the radiologist and with the DL model.

mograms as almost entirely fatty, 48% as scattered areas of fibroglandular tissue, 38% as heterogeneously dense, and 2% as extremely dense, while the original interpreting radiologist assessed 9% as almost entirely fatty, 55% as scattered areas of fibroglandular tissue, 31% as heterogeneously dense, and 5% as extremely dense. The DL model assessed 8% as almost entirely fatty, 57% as scattered areas of fibroglandular tissue, 33% as heterogeneously dense, and 2% as extremely dense. As shown in Table 3, agreement between the DL model and the original interpreting radiologist was good (κ = 0.62; 95% CI: 0.57, 0.67). In addition, agreement between the DL model and the radiologists working in consensus was good (κ = 0.78; 95% CI: 0.73, 0.82), with a significantly higher κ value. Agreement between the radiologists in consensus and the original interpreting radiologist was also good (κ = 0.63; 95% CI: 0.58, 0.69) and closely matched agreement of the DL model with the original interpreting radiologist. Overall, the model followed the assessment of the radiologists in consensus most closely, with the highest level of agreement occurring between the DL model and radiologists in consensus (higher than the agreement between the original interpreting radiologist and the radiologists in consensus).

Clinical Implementation and Acceptance of DL Assessment

A total of 10763 consecutive screening mammograms from 10763 patients were evaluated with the DL model in real time during the clinical implementation phase. The

Table 3: Agreement Statistics across Four BI-RADS Categories in Test, Blinded Reader Consensus, and Clinical Implementation Settings

Setting and Comparison	Linear Weighted κ Value	No. of Mammograms
Test set (DL model vs original interpreting radiologist)	0.67 (0.66, 0.68)	8677
Reader consensus		
DL model vs original interpreting radiologist	0.62 (0.57, 0.67)	500
Reader consensus vs original interpreting radiologist	0.63 (0.58, 0.69)	500
DL model vs reader consensus	0.78 (0.73, 0.82)	500
Clinical implementation (DL model vs final radiologist assessment)	0.85 (0.84, 0.86)	10763

Note.—Data in parentheses are 95% confidence intervals. DL = deep learning. BI-RADS = Breast Imaging Reporting and Data System.

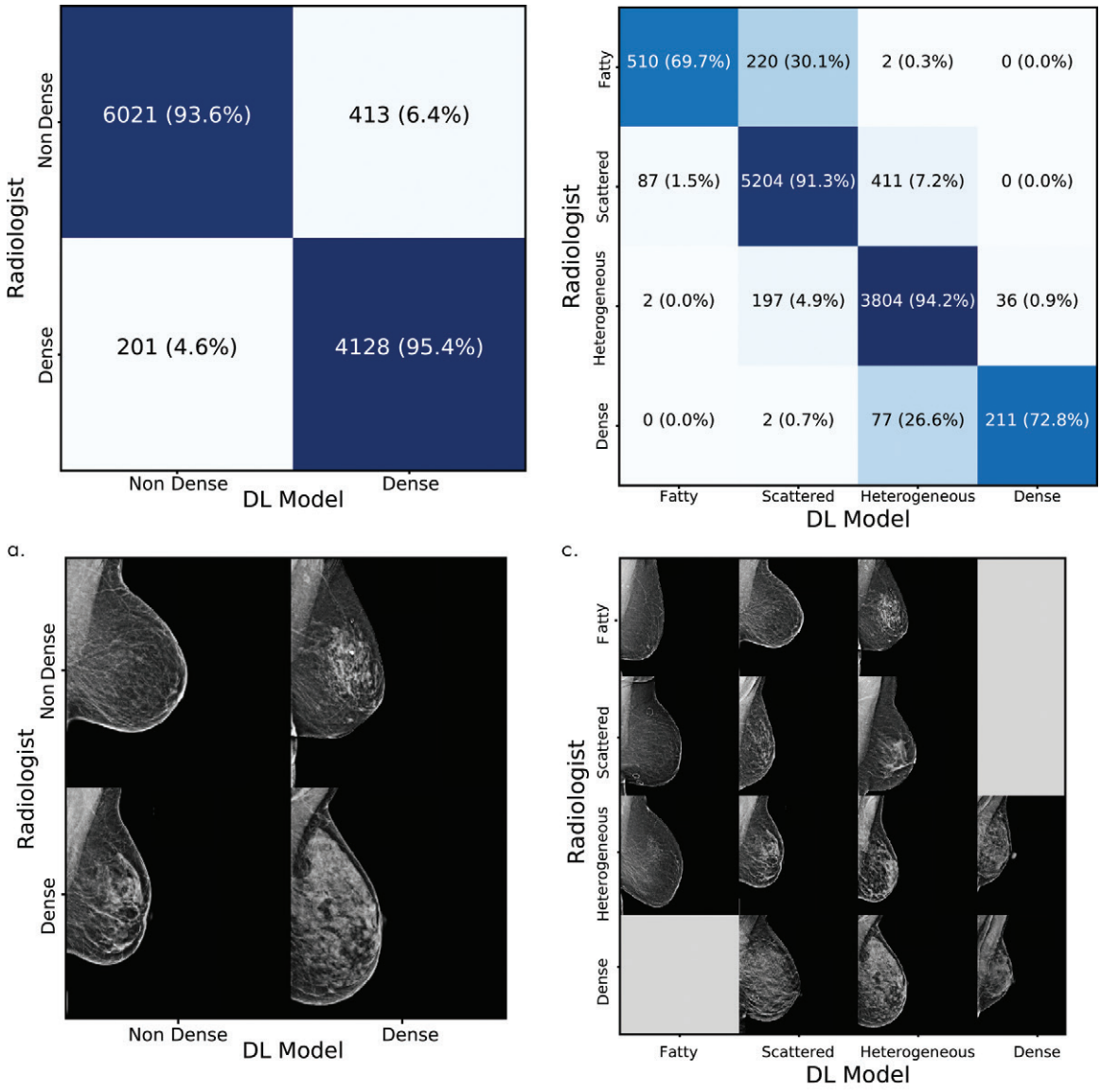

Figure 2: Clinical implementation assessment. Comparison of the original interpreting radiologist assessment with the deep learning (DL) model assessment for **(a)** binary and **(c)** four-way mammographic breast density classification. **(b, d)** Examples of mammograms with concordant and discordant assessments by the radiologist (**b**) and with the DL model (**d**).

DL model assessed 6% of mammograms as almost entirely fatty, 52% as scattered areas of fibroglandular tissue, 40% as heterogeneously dense, and 2% as extremely dense. In the binary categorization of dense or nondense breasts, the proportion of DL assessments accepted by the interpreting radiologist was 10 149 of 10 763 (94%; 95% CI: 94%, 95%) (Table 2). Of the 614 DL assessments not accepted by the interpreting radiologist, 201 (33%) mammograms were downgraded from dense

to nondense, and 413 (67%) were upgraded from nondense to dense (Fig 2). In the four-way BI-RADS categorization, 9729 of 10763 (90%; 95% CI: 90%, 91%) DL assessments were accepted by the interpreting radiologist (Table 2). Agreement between final assessment and DL assessment was very good (κ = 0.85; 95% CI: 0.84, 0.86) (Table 3).

Discussion

Inconsistency in density assessment of mammograms has been widely recognized for the potential to cause patient anxiety and result in unnecessary procedures. To address this issue, we developed a DL model to assess mammographic breast density that was trained by using the assessments of experienced breast imagers. Our DL model was deployed in the mammography clinic to assess performance and acceptance in a large academic breast imaging practice. In this setting, the DL model density assessment was accepted as the final reading in 90% of mammograms by an experienced breast imager.

Most prior methods for automated breast density assessment have not used machine learning or DL models and have shown variable agreement with the density assessments of blinded radiologists. Agreement varied in a retrospective study of 1185 mammograms by Youk et al, with κ scores of 0.54–0.61 for Quantra and 0.32–0.43 for Volpara (11). In a study of 6081 patients, Brandt et al showed moderate agreement, with κ scores of 0.46 for Quantra and 0.57 for Volpara (13). More recently, Wu et al used a DL model to assess density in a reader study of 100 mammograms and showed moderate agreement between their DL model and the assessment of an experienced breast imager, with a κ score of 0.48 (12). In contrast, our DL model showed higher agreement, with κ scores of 0.67 (95% CI: 0.66, 0.68) and 0.78 (95% CI: 0.73, 0.82) in our test set and reader study, respectively. Additionally, our model showed very good agreement (κ = 0.85 [95% CI: 0.84, 0.86]) in our nonblinded clinical implementation setting.

There were limitations to our study. Our reference standard was breast density assessed by the original interpreting radiologist, and this is known to be prone to moderate inter- and intrareader variation. In addition, guidance in BI-RADS fourth and fifth editions for density assessments has changed over time, and this could influence our results. Prior reports have used density assessment at breast MRI or CT as the reference standard to compare with automated mammographic breast density assessment. However, this approach has important limitations, including small sample sizes (range, 100–200 patients) and sample bias toward patients at high risk (21,22). To develop and test our DL model, we used mammograms from more than 30000 women at average risk; it would not have been feasible to incorporate MRI in this setting. Also, this model was trained on mammograms at one academic center that used mammography units from one vendor (Hologic), and further testing on diverse mammograms acquired with machines from multiple vendors and from different institutions is needed. Finally, during the clinical implementation of our project, acceptance of the DL density assessment was measured in an unblinded manner. We acknowledge this "acceptance" metric is not the same as measuring "truth" or "accuracy." However, in a subset of 500 mammograms, the DL assessment showed good agreement (κ = 0.78; 95% CI: 0.73, 0.82) with the consensus interpretation of five experienced blinded breast imagers; this is similar to the very good agreement (κ = 0.85; 95% CI: 0.84, 0.86) between the DL model and the final radiologist assessment.

In summary, we present an analysis of clinical implementation of a DL model used to assess breast density in women undergoing screening digital mammography. Our DL model provides efficient and reliable density assessments, both at the patient level and at the population level, and it is designed to be widely available, simple to use, and cost effective. It can be used to measure breast density in a diverse set of patients, without limitations based on prior surgery or other breast interventions. Our tool can potentially address concerns for current breast density legislation, and it can help providers supply more accurate information to patients and help health systems optimize the use of supplemental screening resources. To this end, we have made our tool publicly available for research use at *http://learningtocure.csail.mit.edu.*

Acknowledgments: The authors are grateful to Christine Edmonds, MD, and Randy C. Miles, MD, MPH, for their participation in the reader study and to Thomas Schultz, BS, Steven Graham, MBA, and Alexander Schultz (Partners Enterprise Medical Imaging [Boston, Mass]) for their support of the clinical implementation phase of this project.

Author contributions: Guarantors of integrity of entire study, C.D.L., A.Y., R.B.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, C.D.L., A.Y., B.D., M.B., R.B.; clinical studies, C.D.L., A.Y., B.D.; statistical analysis, A.Y., T.S., R.B.; and manuscript editing, C.D.L., A.Y., B.D., M.B., R.B.

Disclosures of Conflicts of Interest: **C.D.L.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for GE Healthcare; institution received grants from GE Healthcare. Other relationships: institution has submitted a patent for this density assessment algorithm. **A.Y.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed no relevant relationships. Other relationships: institution has submitted a patent for this density assessment algorithm. **T.S.** disclosed no relevant relationships. **B.D.** disclosed no relevant relationships. **M.B.** disclosed no relevant relationships. **K.S.** disclosed no relevant relationships. **R.B.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is on the scientific advisory board to Janssen; is on the advisory boards of Meltwater and Asapp; gave lectures for Nuance. Other relationships: institution has submitted a patent for this density assessment algorithm.

References

1. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 1995;87(9):670–675.
2. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168–175.
3. Whitehead J, Carlile T, Kopecky KJ, et al. Wolfe mammographic parenchymal patterns. a study of the masking hypothesis of Egan and Mosteller. Cancer 1985;56(6):1280–1286.
4. 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.
5. Bahl M, Baker JA, Bhargavan-Chatfield M, Brandt EK, Ghate SV. Impact of breast density notification legislation on radiologists' practices of reporting breast density: a multi-state study. Radiology 2016;280(3):701–706.

6. Spayne MC, Gard CC, Skelly J, Miglioretti DL, Vacek PM, Geller BM. Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. Breast J 2012;18(4):326–333.
7. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol 2000;174(6):1769–1777.
8. Ray KM, Price ER, Joe BN. Breast density legislation: mandatory disclosure to patients, alternative screening, billing, reimbursement. AJR Am J Roentgenol 2015;204(2):257–260.
9. Sobotka J, Hinrichs C. Breast density legislation: discussion of patient utilization and subsequent direct financial ramifications for insurance providers. J Am Coll Radiol 2015;12(10):1011–1015.
10. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst 1998;90(23):1801–1809.
11. Youk JH, Gweon HM, Son EJ, Kim JA. Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: a comparison with visual assessment. AJR Am J Roentgenol 2016;206(5):1056–1062.
12. Wu N, Geras KJ, Shen Y, et al. 2017. Breast density classification with deep convolutional neural networks. arXiv preprint arXiv:1711.03674. https://arxiv.org/abs/1711.03674. Accessed August 8, 2018.
13. Brandt KR, Scott CG, Ma L, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiology 2016;279(3):710–719.
14. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. Radiology 2018;286(3):810–818.
15. 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.
16. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017;284(2):574–582.
17. Geras KJ, Wolfson S, Shen Y, et al. 2017. High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047. https://arxiv.org/abs/1703.07047. Accessed August 8, 2018.
18. Kallenberg M, Petersen K, Nielsen M, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 2016;35(5):1322–1331.
19. American College of Radiology. ACR BI-RADS Atlas—mammography. 5th ed. Reston, Va: American College of Radiology, 2013.
20. He K, Zhang X, Ren S, Sun J. 2015. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385. https://arxiv.org/abs/1512.03385. Accessed August 8, 2018.
21. Gubern-Mérida A, Kallenberg M, Platel B, Mann RM, Martí R, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms: a validation study. PLoS One 2014;9(1):e85952.
22. Wang J, Azziz A, Fan B, et al. Agreement of mammographic measures of volumetric breast density to MRI. PLoS One 2013;8(12):e81653.