



Published in final edited form as:

Ann Surg Oncol. 2023 November ; 30(12): 7107–7115. doi:10.1245/s10434-023-14083-1.

Analysis of Specimen Mammography with Artificial Intelligence to Predict Margin Status

Kevin A Chen, MD¹, Kathryn E Kirchoff, BS², Logan R Butler, MD¹, Alexa D Holloway, BS¹, Muneera R Kapadia, MD¹, Cherie M Kuzmiak, DO³, Stephanie M Downs-Canner, MD⁴, Phillip M Spanheimer, MD¹, Kristalyn K Gallagher, DO^{1,*}, Shawn M Gomez, EngScD^{2,*}

¹Department of Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC

²Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC

³Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁴Department of Surgery, Breast Service, Memorial Sloan Kettering Cancer Center, New York, NY

Abstract

Background: Intra-operative specimen mammography is a valuable tool in breast cancer surgery, providing immediate assessment of margins for a resected tumor. However, the accuracy of specimen mammography in detecting microscopic margin positivity is low. We sought to develop an artificial intelligence model to predict the pathologic margin status of resected breast tumors using specimen mammography.

Methods: A dataset of specimen mammography images matched with pathologic margin status was collected from our institution from 2017-2020. The dataset was randomly split into training, validation, and test sets. Specimen mammography models pre-trained on radiologic images were developed and compared with models pre-trained on non-medical images. Model performance was assessed using sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC).

Results: The dataset included 821 images and 53% had positive margins. For three out of four model architectures tested, models pre-trained on radiologic images outperformed non-medical models. The highest performing model, InceptionV3, showed a sensitivity of 84%, a specificity of 42%, and AUROC of 0.71. Model performance was better among patients with invasive cancers, less dense breasts, and non-White race.

Conclusions: This study developed and internally validated artificial intelligence models which predict pathologic margins status for partial mastectomy from specimen mammograms. The models' accuracy compares favorably with the published literature on surgeon and radiologist interpretation of specimen mammography. With further development, these models could more precisely guide the extent of resection, potentially improving cosmesis and reducing re-operations.

*Co-corresponding authors Kristalyn K. Gallagher, Division of Surgical Oncology, Department of Surgery, University of North Carolina at Chapel Hill, Burnett-Womack Building, CB#7213, Chapel Hill, NC 27599-7213, kristalyn_gallagher@med.unc.edu; Shawn M. Gomez, Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill, 10202C Mary Ellen Jones Building, CB#7575, Chapel Hill, NC 27599-7575, smgomez@unc.edu.

Introduction

Breast-conserving surgery with radiation is often the preferred approach for early-stage breast cancer, balancing oncologic resection and cosmesis.¹ In this approach, obtaining negative margins is critical for reducing local recurrence. However, reported re-operation rates for positive margins remain high, at about 15%.²⁻⁶ Multiple interventions have attempted to identify positive margins intra-operatively, including specimen mammography.

Specimen mammography is a widely used technique that provides immediate feedback on the quality of resection and may assist surgeons with identifying suspicious margins and directing targeted removal of additional tissue.⁷ However, specimen mammography can be inaccurate, with sensitivity ranging from 20% to 58% and area under the receiver operating characteristic curve (AUROC) of ~0.7.⁸⁻¹¹ A tool to assist clinicians with interpretation of specimen mammography for partial mastectomy would be beneficial for improving identification of positive margins and reducing rates of re-operation. This tool would be particularly useful for low-volume and low-resource centers, which see higher rates of positive margins.¹²

In addition, in part because of the limitations of specimen mammography and other intra-operative margin assessment techniques, universal use of cavity shave margins has been adopted as a highly effective means of reducing positive margins.¹³ However, this technique does result in a higher volume of tissue resected and higher costs for pathology.^{14,15} An automated method of assessing the primary specimen could assist surgeons with selective use of cavity shave margins.

Deep learning, commonly referred to as artificial intelligence (AI), is an emerging field that has enabled computational interpretation of medical images.¹⁶ It has been extensively applied to screening mammography, with a recent systematic review identifying 82 relevant studies describing high accuracy for identifying breast cancer.¹⁷ Deep learning also shows promise for intra-operative use, with recent applications to the interpretation of laparoscopic video.¹⁸⁻²⁰ However, deep learning has not yet been applied to intra-operative specimen mammography.

The goal of this study is to develop a computer vision model that can predict margin status for partial mastectomy based on specimen mammography alone. We hypothesized that this model could demonstrate clinically actionable accuracy and outperform human accuracy in predicting pathologic margin status.

Methods

Dataset

Prior to data collection, approval from the University of North Carolina Institutional Review Board (#20-1820) was obtained. We included all consecutive patients undergoing partial mastectomy for breast cancer, including invasive lobular carcinoma, invasive ductal carcinoma, and ductal carcinoma in situ (DCIS), for whom specimen mammography and pathologic margin status were available. Data was collected in two phases. First,

from 7/2017 to 6/2020, specimen mammograms were prospectively collected as part of a single surgeon's quality assurance process. Second, from 8/2020 to 4/2022, specimen mammograms from four surgeons were collected specifically for this project. In the second phase of data collection, chart review of pre-operative mammography and pathology records was used to obtain additional clinical information including age, race/ethnicity, BMI, breast density,²¹ tumor type, and tumor grade.

A single 2D, anterior-posterior image was selected for each specimen. All images were obtained using a Faxitron[®] Trident[®] HD Specimen Radiography System (Hologic, Marlborough, Massachusetts), intra-operatively, immediately after resection of the specimen. Our institution's process for margin processing/analysis is to obtain a specimen mammogram, which is interpreted intra-operatively by a breast radiologist, routinely obtain cavity shave margins, and submit the primary specimen and margins for formal pathology.

Primary outcome

Using retrospective chart review, specimen mammograms were matched with pathology reports and categorized into positive and negative classes based on National Comprehensive Cancer Network (NCCN) guidelines.²² For specimens with invasive cancer, a positive margin was defined as "ink on tumor." For specimens with DCIS, a positive margin was defined as DCIS within 2mm of the margin. For specimens containing both DCIS and invasive cancer, but with DCIS within 2mm of the margin or mixed pathology, we elected to categorize this pathology as positive, acknowledging that, clinically, this result is treated as negative. We used this approach for model training to maximize the sensitivity of the model for DCIS margins and avoid training the model to ignore DCIS in specimens with invasive cancer. In addition, we used specimen mammograms and pathologic margin status of the main specimen, excluding cavity shave margins, which were routinely used, but not routinely imaged.

Data processing

The dataset was divided randomly into training, validation, and test sets in a 60/20/20 ratio. A single anterior-posterior image was used for each specimen and resized to 512 x 512 pixels. The training set underwent standard data augmentation including random flipping (vertically and horizontally), zoom, shifts (horizontal and vertical), and rotation.²³

Pre-training Datasets

We compared two different pre-training strategies. First, we used models from the RadImageNet project, which developed pre-trained models based on 1.35 million annotated medical images.²⁴ In contrast, prior medical computer vision projects have often relied on ImageNet, which is a large database of >14 million images. However, ImageNet includes largely images of non-medical objects, such as balloons or strawberries.²⁵ We compared the performance of "radiology-specific models," pre-trained on RadImageNet, with "general models," pre-trained on ImageNet.

Modeling

We developed models based on the architectures used by the RadImageNet project, including ResNet-50, InceptionV3, Inception-ResNet-v2, and DenseNet-121.²⁴ These are different types of convolutional neural networks, a type of AI model specifically used for computer vision. To this end, we implemented eight models: one of each base architecture pre-trained on RadImageNet and one of each pre-trained on ImageNet. After the highest performing model type and pre-training strategy was identified, the model was further optimized. Additional details on model development are available in the GitHub repository referenced at the end of this section. A diagram of the model is shown in Figure 1a.

Model evaluation

Model performance for predicting pathologic margin status was assessed primarily using area under the receiver operating curve (AUROC). This is a classification metric that assesses a model's ability to distinguish positive from negative cases and ranges from 0.5 at the worst, to 1 at the best. For the highest performing model (based on AUROC), evaluation metrics were calculated for categories within tumor type, breast density, and race/ethnicity because screening mammography has previously been shown to have different accuracy within these groups.^{26,27} For this analysis, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the precision-recall curve (AUPRC) were also calculated. In addition, Grad-CAM was used to create saliency maps to assess pixel importance for 10 randomly selected images from the test set and these were visually assessed.²⁸ Figure 1b shows the overall workflow for model development and evaluation.

Models were implemented and evaluated using the Python (version 3.8) libraries scikit-learn and Tensorflow/Keras.^{29,30} An NVIDIA RTX A4500 graphics card (NVIDIA, Santa Clara, CA) was used for model training and validation. For characterization of the cohort, Chi-squared test was used to compare categorical variables and T-test to compare continuous variables, using the tableone package.³¹ Code to reproduce this work and additional details on methods are available at github.com/gomezlab/cvsm.

Results

The dataset included 806 images. Of these, 450 were collected in the first, single surgeon phase, while 356 were collected in the second, multiple surgeon phase. 431 (52.5%) images had positive margins in the main specimen without considering cavity shave margins. Representative images for each classification are shown in Figure 3. Within the positive margin classification, 128 had mixed pathology with both invasive cancer and DCIS 2mm from the margin. The average age was 60 and most patients had IDC (70.0%) or DCIS (21.3%). Non-Hispanic White patients comprised 70.1% of the cohort, compared with 18.1% Non-Hispanic Black and 7.8% Hispanic. Most patients had a breast density of B (scattered fibroglandular densities) (45.8%) or C (heterogeneously dense) (36.9%). Patients with infiltrating ductal carcinoma who had a tumor grade of 2 were more likely to have positive margins (Table 1). After splitting into training, validation, and testing groups, 485 images were used for training, 160 for validation, and 161 for testing.

We first compared the performance of models pre-trained on radiologic images (RadImageNet), compared with models pre-trained on general images (ImageNet). Overall, radiology-specific models showed higher accuracy compared with general models with AUROC 0.63 to 0.71 vs. 0.46 – 0.68, respectively. Each of the model types tested showed higher performance with radiology-specific pre-training, except for one (DenseNet121), and performance was similar between the two pre-training strategies in this case (AUROC 0.66 vs 0.68). The highest performing model was InceptionV3 with radiology-specific pre-training. Receiver operating characteristic and precision-recall curves are shown for all models in Figure 4.

Based on this analysis, InceptionV3 with radiology-specific pre-training was further optimized. This model showed an AUROC of 0.71, AUPRC of 0.73, sensitivity of 85%, specificity of 45%, positive predictive value of 62%, and negative predictive value of 70%. In subset analysis, we found that model performance was worse for DCIS (AUROC 0.65) compared with invasive ductal carcinoma (AUROC 0.71) and invasive lobular carcinoma (AUROC 0.75). We also found that model performance was worse for extremely dense breast tissue (category D) compared with less dense breast tissue (Table 2). For race/ethnicity, model performance was worse for Non-Hispanic White patients compared with non-White patients, although this difference appeared to be partially driven by breast density, with 14% of Non-Hispanic White patients having category D breast density, compared with 9% of non-White patients (Table 3).

To improve model interpretability, or understanding what parts of the image contributed the most to model decision-making, we assessed where the attention of the model was focused. In most cases, models pre-trained on radiologic images had attention focused on relevant parts of the image, such as localization wires, biopsy clips, and areas of tumor, while models pre-trained on non-medical images did not (Figure 5).

Discussion

This project developed an artificial intelligence model that predicts the pathologic margin status of partial mastectomy specimens based on specimen mammograms. Pre-training with radiology images was found to improve model predictions compared with pre-training with non-medical images. With an internal test set, the model showed a sensitivity of 85%, a specificity of 45%, and AUROC of 0.71. Analysis of pixel importance suggested that model attention was focused on relevant portions of the image.

Despite advances in intra-operative margin assessment, the rate of positive margins after partial mastectomy remains high.³ In fact, due to limitations in visual, tactile, and radiographic intraoperative assessment of margins, cavity shave margins have been widely adopted to reduce the rate of positive margins.^{13,14} Still, specimen mammography is a widely used method that benefits from its availability within the operating room and its ability to provide immediate feedback. Because of this, specimen imaging for non-palpable lesions is recommended by the Collaborative Attempt to Lower Lumpectomy Reoperation rates (CALLER) toolbox.³² However, the diagnostic accuracy of specimen mammography is highly variable and relatively low, with reported sensitivity ranging from 20-58%.⁸⁻¹¹ A

meta-analysis of nine studies showed a pooled sensitivity of 53% (95% CI 45 – 61%) with a pooled specificity of 84% (95% CI 77 - 89%).³³ AUROC is similarly variable, but low, ranging from 0.60 to 0.73.³³⁻³⁷ Our models' accuracy metrics are higher than many previous results published in the literature, demonstrating the potential of this approach.

Another interesting finding from our study was that our model showed lower accuracy in predicting margin status among patients with the highest breast density (category D). This is expected, given similar findings for screening mammography, but has not been previously reported for specimen mammography.²⁷ In contrast to previous studies applying artificial intelligence (AI) to mammography, our models show higher accuracy among non-White patients compared with White patients.²⁶ This difference may be because non-White patients are well-represented within our dataset and because of the higher percentage of category D breast density among White patients in this study. In addition, our study agrees with previous literature showing that DCIS margins are particularly difficult to assess using specimen mammography.^{7,38} Use of a larger training set focused on DCIS alone may be necessary to overcome this limitation.

More generally, the overall performance of our models agrees with other recent advances in computer vision classification of mammograms that show improved accuracy when radiologists are assisted by AI.³⁹⁻⁴¹ Recently, AI has also been successfully applied to laparoscopic video, demonstrating its potential to assist with real-time, intra-operative decision-making.⁴²⁻⁴⁴ AI-assisted interpretation of specimen mammography may function similarly. Low-volume or low-resource centers that lack access to dedicated breast surgeons or radiologists have the highest rates of positive margins and may benefit most from AI-assistance, raising the potential for these systems to improve equity in surgical outcomes.^{12,45} AI-assisted clinical decision-support systems would be most useful to these clinicians and thus have the potential to improve equity in surgical outcomes for their patients.⁴⁶

There are two clinical scenarios that would specifically benefit from our models. First, if a positive margin was predicted, this could guide additional resection. Alternatively, identifying negative margins on the primary specimen could allow precise patient-specific omission of cavity shave margins which could decrease the overall volume of resection, improve cosmesis, and decrease costs related to pathology.¹⁵ Because of cavity shave margins, we were not able to develop a model for margin positivity based on the final margin, so we focused on the second scenario, confidently identifying negative margins. To maximize the negative predictive value, we trained the model on the widest margin guideline, DCIS. We chose this approach to maximize the negative predictive value, and because false positives in this setting are less problematic as the patient would simply receive routine cavity shave margins. The threshold of NPV for clinical use is an open question and dependent on individual surgeon judgment. However, further development of these models is warranted, including collection of a larger dataset to ensure reliability across different imaging hardware and institutions and mammography-specific transfer learning to improve model accuracy.

This project has several limitations. First, the accuracy and generalizability of the model is most significantly limited by the size and single-institution nature of the dataset. A larger, multi-institutional dataset would likely result in a more accurate and robust model and could verify its external validity. Second, the rate of positive margins is higher than previously reported because of our exclusion of cavity shave margins and use of DCIS margins in mixed invasive cancer/DCIS pathology.³ This approach maximizes the sensitivity of the model for DCIS margins, reduces false negatives, and results in a single model which can be used for all cases. Third, our model does not identify which side of the specimen may have a positive margin. Leveraging model attention techniques to automatically localize image features associated with a positive margin is possible and represents a direction for future research.^{47,48} Finally, we did not assess radiologist or surgeon accuracy for margin prediction on our dataset, although this is likely to be similar to previous literature.

More recent imaging techniques, such as 3D tomosynthesis or optical coherence tomography, may also be more accurate compared with 2D specimen mammography and result in improved models.⁴⁹⁻⁵¹ In addition, trials of new technologies for margin assessment such as fluorescence and spectroscopy have shown promising results.⁵²⁻⁵⁷ However, these devices require purchase of costly hardware and carry the risk of allergic reaction. In contrast, the current approach has the advantage of using widely available systems.

Conclusion

In conclusion, we developed a prototype model that predicts the pathologic margin status of partial mastectomy specimens based on intra-operative specimen mammography. The model's predictions compare highly favorably with human interpretation. Optimized and externally validated versions of this model could assist surgeons with predicting margin status intra-operatively, guide selective use of cavity shave margins, and ultimately reduce the need for re-operation in breast-conserving surgery.

Acknowledgements

Disclosures: The authors KAC, KKG and SMG hold a preliminary patent describing the methods used in this study. This work was supported by funding from the National Institutes of Health (Program in Translational Medicine T32-CA244125 to UNC/KAC).

References

1. Fisher B, Anderson S, Bryant J, et al. Twenty-Year Follow-up of a Randomized Trial Comparing Total Mastectomy, Lumpectomy, and Lumpectomy plus Irradiation for the Treatment of Invasive Breast Cancer. *N Engl J Med.* 2002;347(16):1233–1241. doi:10.1056/nejmoa022152 [PubMed: 12393820]
2. Marinovich ML, Azizi L, Macaskill P, et al. The Association of Surgical Margins and Local Recurrence in Women with Ductal Carcinoma In Situ Treated with Breast-Conserving Therapy: A Meta-Analysis. *Ann Surg Oncol.* 2016;23(12):3811–3821. doi:10.1245/s10434-016-5446-2 [PubMed: 27527715]
3. Morrow M, Abrahamse P, Hofer TP, et al. Trends in Reoperation After Initial Lumpectomy for Breast Cancer Addressing Overtreatment in Surgical Management Author Audio Interview Supplemental content. *JAMA Oncol.* 2017;3(10):1352–1357. doi:10.1001/jamaoncol.2017.0774 [PubMed: 28586788]

4. Moran MS, Schnitt SJ, Giuliano AE, et al. SSO-ASTRO Consensus Guideline on Margins for Breast-Conserving Surgery with Whole Breast Irradiation in Stage I and II Invasive Breast Cancer. *Int J Radiat Oncol Biol Phys*. 2014;88(3):553. doi:10.1016/j.IJROBP.2013.11.012 [PubMed: 24521674]
5. Houssami N, Macaskill P, Luke Marinovich M, Morrow M. The association of surgical margins and local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy: A meta-analysis. *Ann Surg Oncol*. 2014;21(3):717–730. doi:10.1245/s10434-014-3480-5 [PubMed: 24473640]
6. Rosenberger LH, Mamtani A, Fuzesi S, et al. Early Adoption of the SSO-ASTRO Consensus Guidelines on Margins for Breast-Conserving Surgery with Whole-Breast Irradiation in Stage I and II Invasive Breast Cancer: Initial Experience from Memorial Sloan Kettering Cancer Center. *Ann Surg Oncol*. 2016;23(10):3239–3246. doi:10.1245/S10434-016-5397-7 [PubMed: 27411549]
7. Versteegden DPA, Keizer LGG, Schlooz-Vries MS, Duijm LEM, Wauters CAP, Strobbe LJA. Performance characteristics of specimen radiography for margin assessment for ductal carcinoma in situ: a systematic review. *Breast Cancer Res Treat*. 2017;166(3):669–679. doi:10.1007/s10549-017-4475-2 [PubMed: 28831674]
8. Layfield DM, May DJ, Cutress RI, et al. The effect of introducing an in-theatre intra-operative specimen radiography (IOSR) system on the management of palpable breast cancer within a single unit. *Breast*. 2012;21(4):459–463. doi:10.1016/j.breast.2011.10.010 [PubMed: 22119488]
9. Bathla L, Harris A, Davey M, Sharma P, Silva E. High resolution intra-operative two-dimensional specimen mammography and its impact on second operation for re-excision of positive margins at final pathology after breast conservation surgery. *Am J Surg*. 2011;202:387–394. doi:10.1016/j.amjsurg.2010.09.031 [PubMed: 21943945]
10. Hisada T, Sawaki M, Ishiguro J, et al. Impact of intraoperative specimen mammography on margins in breast-conserving surgery. *Mol Clin Oncol*. 2016;5(3):269–272. doi:10.3892/mco.2016.948 [PubMed: 27588192]
11. Funk A, Heil J, Harcos A, et al. Efficacy of intraoperative specimen radiography as margin assessment tool in breast conserving surgery. *Breast Cancer Res Treat*. 2020;179(2):425–433. doi:10.1007/s10549-019-05476-6 [PubMed: 31654190]
12. Yen TWF, Pezzin LE, Li J, Sparapani R, Laud PW, Nattinger AB. Effect of hospital volume on processes of breast cancer care: A National Cancer Data Base study. *Cancer*. 2017;123(6):957–966. doi:10.1002/CNCR.30413 [PubMed: 27861746]
13. Chagpar AB, Killelea BK, Tsangaris TN, et al. A randomized, controlled trial of cavity shave margins in breast cancer. *N Engl J Med*. 2015;373(6):503–510. doi:10.1056/NEJMoa1504473 [PubMed: 26028131]
14. Dupont E, Tsangaris T, Garcia-Cantu C, et al. Resection of Cavity Shave Margins in Stage 0–III Breast Cancer Patients Undergoing Breast Conserving Surgery. *Ann Surg*. 2019;XX(Xx):1. doi:10.1097/sla.0000000000003449
15. Cartagena LC, McGuire K, Zot P, Pillappa R, Idowu M, Robila V. Breast-Conserving Surgeries With and Without Cavity Shave Margins Have Different Re-excision Rates and Associated Overall Cost: Institutional and Patient-Driven Decisions for Its Utilization. *Clin Breast Cancer*. 2021;21(5):e594–e601. doi:10.1016/J.CLBC.2021.03.003 [PubMed: 33814286]
16. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. doi:10.1038/nature14539 [PubMed: 26017442]
17. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *npj Digit Med* 2021 41. 2021;4(1):1–23. doi:10.1038/s41746-021-00438-z
18. 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. doi:10.1148/radiol.2019182716 [PubMed: 31063083]
19. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: Clinical implementation. *Radiology*. 2019;290(1):52–58. doi:10.1148/radiol.2018180694 [PubMed: 30325282]

20. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann Surg.* 2019;270(3):414–421. doi:10.1097/SLA.0000000000003460 [PubMed: 31274652]
21. Radiology AC of, D'Orsi CJ. ACR BI-RADS Atlas: Breast Imaging Reporting and Data System: 2013. Published online 2018. Accessed May 3, 2023. https://books.google.com/books/about/2013_ACR_BI_RADS_Atlas.html?id=nhWSjwEACAAJ
22. Gradishar WJ, Anderson BO, Abraham J, et al. Breast cancer, version 3.2020. *JNCCN J Natl Compr Cancer Netw.* 2020;18(4):452–478. doi:10.6004/jnccn.2020.0016
23. Zheng A, Casari A. Feature Engineering for Machine Learning PRINCIPLES AND TECHNIQUES FOR DATA SCIENTISTS.
24. Mei X, Liu Z, Robson PM, et al. RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. *Radiol Artif Intell.* 2022;4(5). doi:10.1148/ryai.210315
25. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis.* 2015;115(3):211–252. doi:10.1007/s11263-015-0816-y
26. Yala A, Mikhael PG, Strand F, et al. Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model. *J Clin Oncol.* 2022;40(16):1732–1740. doi:10.1200/JCO.21.01337 [PubMed: 34767469]
27. Siu AL. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med.* 2016;164(4):279–296. doi:10.7326/M15-2886 [PubMed: 26757170]
28. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int J Comput Vis.* 2016;128(2):336–359. doi:10.1007/s11263-019-01228-7
29. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
30. Chollet F, others. Keras. Published online 2015. <https://github.com/fchollet/keras>
31. Pollard TJ, Johnson AEW, Raffa JD, Mark RG. tableone: An open source Python package for producing summary statistics for research papers. *JAMIA Open.* 2018;1(1):26. doi:10.1093/JAMIAOPEN/OOY012 [PubMed: 31984317]
32. Landercasper J, Attai D, Atisha D, et al. Toolbox to Reduce Lumpectomy Reoperations and Improve Cosmetic Outcome in Breast Cancer Patients: The American Society of Breast Surgeons Consensus Conference. *Ann Surg Oncol.* 2015;22(10):3174. doi:10.1245/S10434-015-4759-X [PubMed: 26215198]
33. St John ER, Al-Khudairi R, Ashrafian H, et al. Diagnostic accuracy of intraoperative techniques for margin assessment in breast cancer surgery a meta-analysis. *Ann Surg.* 2017;265(2):300–310. doi:10.1097/SLA.0000000000001897 [PubMed: 27429028]
34. Kulkarni SA, Kulkarni K, Schacht D, et al. High-Resolution Full-3D Specimen Imaging for Lumpectomy Margin Assessment in Breast Cancer. *Ann Surg Oncol.* 2021;28(10):5513–5524. doi:10.1245/s10434-021-10499-9 [PubMed: 34333705]
35. Mazouni C, Rouzier R, Balleyguier C, et al. Specimen radiography as predictor of resection margin status in non-palpable breast lesions. *Clin Radiol.* 2006;61(9):789–796. doi:10.1016/j.crad.2006.04.017 [PubMed: 16905388]
36. Mario J, Venkataraman S, Fein-Zachary V, Knox M, Brook A, Slanetz P. Lumpectomy Specimen Radiography: Does Orientation or 3-Dimensional Tomosynthesis Improve Margin Assessment? *Can Assoc Radiol J.* 2019;70(3):282–291. doi:10.1016/j.carj.2019.03.005 [PubMed: 31300313]
37. Manhoobi IP, Bodilsen A, Nijkamp J, et al. Diagnostic accuracy of radiography, digital breast tomosynthesis, micro-CT and ultrasound for margin assessment during breast surgery: A systematic review and meta-analysis. *Acad Radiol.* 2022;29(10):1560–1572. doi:10.1016/J.ACRA.2021.12.006 [PubMed: 34996687]
38. Lange M, Reimer T, Hartmann S, Glass, Stachs A. The role of specimen radiography in breast-conserving therapy of ductal carcinoma in situ. *Breast.* 2016;26:73–79. doi:10.1016/j.breast.2015.12.014 [PubMed: 27017245]
39. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Heal.* 2020;2(3):e138–e148. doi:10.1016/S2589-7500(20)30003-0

40. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J Natl Cancer Inst.* 2019;111(9):916–922. doi:10.1093/JNCI/DJY222 [PubMed: 30834436]
41. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology.* 2019;290(3):305–314. doi:10.1148/radiol.2018181371 [PubMed: 30457482]
42. Madani A, Namazi B, Altieri MS, et al. Artificial Intelligence for Intraoperative Guidance: Using Semantic Segmentation to Identify Surgical Anatomy during Laparoscopic Cholecystectomy. *Ann Surg.* 2022;276(2):363–369. doi:10.1097/SLA.0000000000004594 [PubMed: 33196488]
43. D K, N T, H MMMM, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc.* 2020;1(11):3. doi:10.1007/s00464-019-07281-0
44. Cheng K, You J, Wu S, et al. Artificial intelligence-based automated laparoscopic cholecystectomy surgical phase recognition and analysis. *Surg Endosc.* 2022;36(5):3160–3168. doi:10.1007/s00464-021-08619-3 [PubMed: 34231066]
45. Istasy P, Lee WS, Iansavichene A, et al. The Impact of Artificial Intelligence on Health Equity in Oncology: Scoping Review. *J Med Internet Res.* 2022;24(11):e39748. doi:10.2196/39748 [PubMed: 36005841]
46. Istasy P, Lee WS, Iansavitchene A, et al. The Impact of Artificial Intelligence on Health Equity in Oncology: A Scoping Review. *Blood.* 2021;138(Supplement 1):4934. doi:10.1182/BLOOD-2021-149264
47. Ghorbani A, Wexler Google Brain J, Zou J, Kim Google Brain B. Towards Automatic Concept-Based Explanations. <https://github.com/amiratag/ACE>
48. Sercan˘ S, Arık S, Pfister T. Protoattend: Attention-Based Prototypical Learning.
49. Partain N, Calvo C, Mokdad A, et al. Differences in Re-excision Rates for Breast-Conserving Surgery Using Intraoperative 2D Versus 3D Tomosynthesis Specimen Radiograph. *Ann Surg Oncol.* 2020;27(12):4767–4776. doi:10.1245/s10434-020-08877-w [PubMed: 32740738]
50. Un Park K, Kuerer HM, Rauch GM, et al. Digital Breast Tomosynthesis for Intraoperative Margin Assessment during Breast-Conserving Surgery. Published online 2019. doi:10.1245/s10434-019-07226-w
51. Polat YD, Ta kın F, Çilda MB, Tanyeri A, Soyder A, Ergin F. The role of tomosynthesis in intraoperative specimen evaluation. *Breast J.* 2018;24(6):992–996. doi:10.1111/TBJ.13070 [PubMed: 29781228]
52. Sandor MF, Schwabach B, Hofmann V, et al. Imaging of lumpectomy surface with large field-of-view confocal laser scanning microscope for intraoperative margin assessment - POLARHIS study. *Breast.* 2022;66:118–125. doi:10.1016/J.BREAST.2022.10.003 [PubMed: 36240525]
53. Aref MH, El-Gohary M, Elrewainy A, et al. Emerging technology for intraoperative margin assessment and post-operative tissue diagnosis for breast-conserving surgery. *Photodiagnosis Photodyn Ther.* 2023;42. doi:10.1016/J.PDPDT.2023.103507
54. Mondal S, Sthanikam Y, Kumar A, et al. Mass Spectrometry Imaging of Lumpectomy Specimens Deciphers Diacylglycerols as Potent Biomarkers for the Diagnosis of Breast Cancer. *Anal Chem.* 2023;95(20). doi:10.1021/ACS.ANALCHEM.3C01019
55. Wang J, Zhang L, Pan Z. Evaluating the impact of radiofrequency spectroscopy on reducing reoperations after breast conserving surgery: A meta-analysis. *Thorac cancer.* 2023;14(16). doi:10.1111/1759-7714.14890
56. Zúñiga WC, Jones V, Anderson SM, et al. Raman Spectroscopy for Rapid Evaluation of Surgical Margins during Breast Cancer Lumpectomy. *Sci Rep.* 2019;9(1). doi:10.1038/S41598-019-51112-0
57. Schnabel F, Boolbol SK, Gittleman M, et al. A randomized prospective study of lumpectomy margin assessment with use of marginprobe in patients with nonpalpable breast malignancies. *Ann Surg Oncol.* 2014;21(5):1589–1595. doi:10.1245/s10434-014-3602-0 [PubMed: 24595800]

Synopsis

Specimen mammography can assess specimen adequacy following partial mastectomy for breast cancer. This study developed and internally validated an artificial intelligence model to predict pathologic margin status from specimen mammograms, with comparable accuracy to surgeons and radiologists.

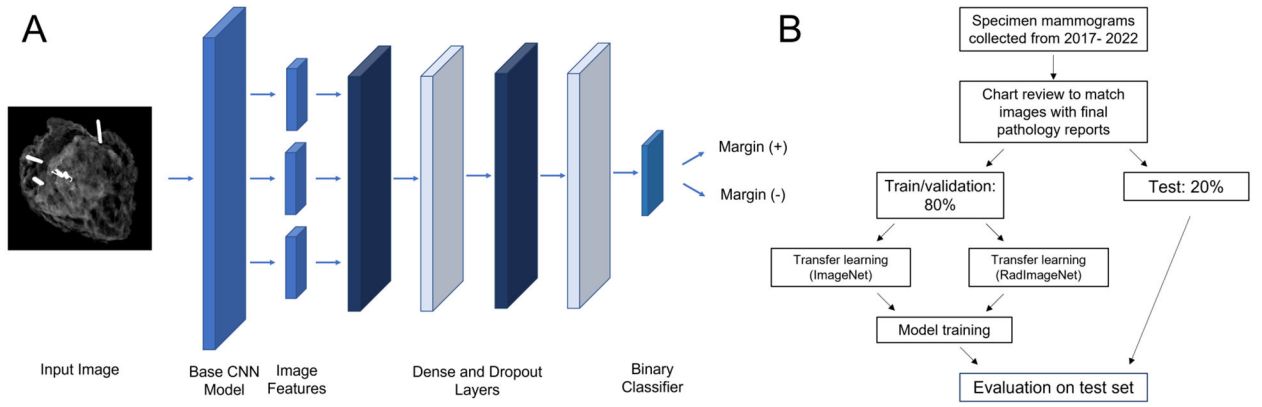


Figure 1.
A - Diagram of model structure, B - Flowchart of study design

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

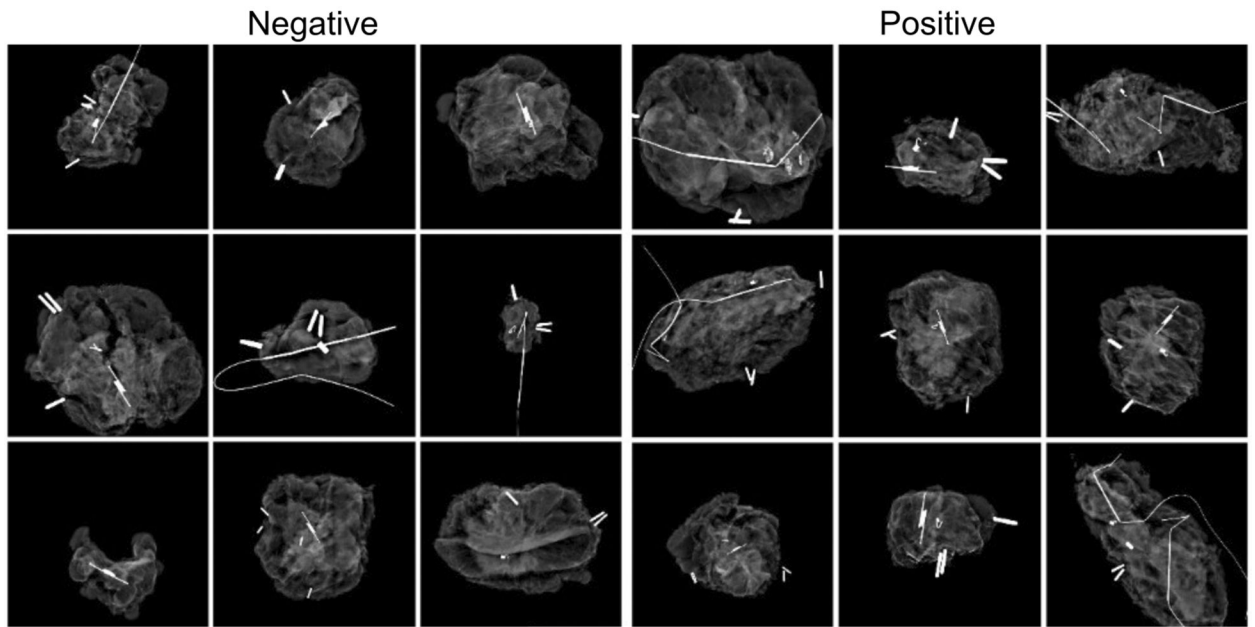


Figure 2.
Example specimen mammograms by margin status classification

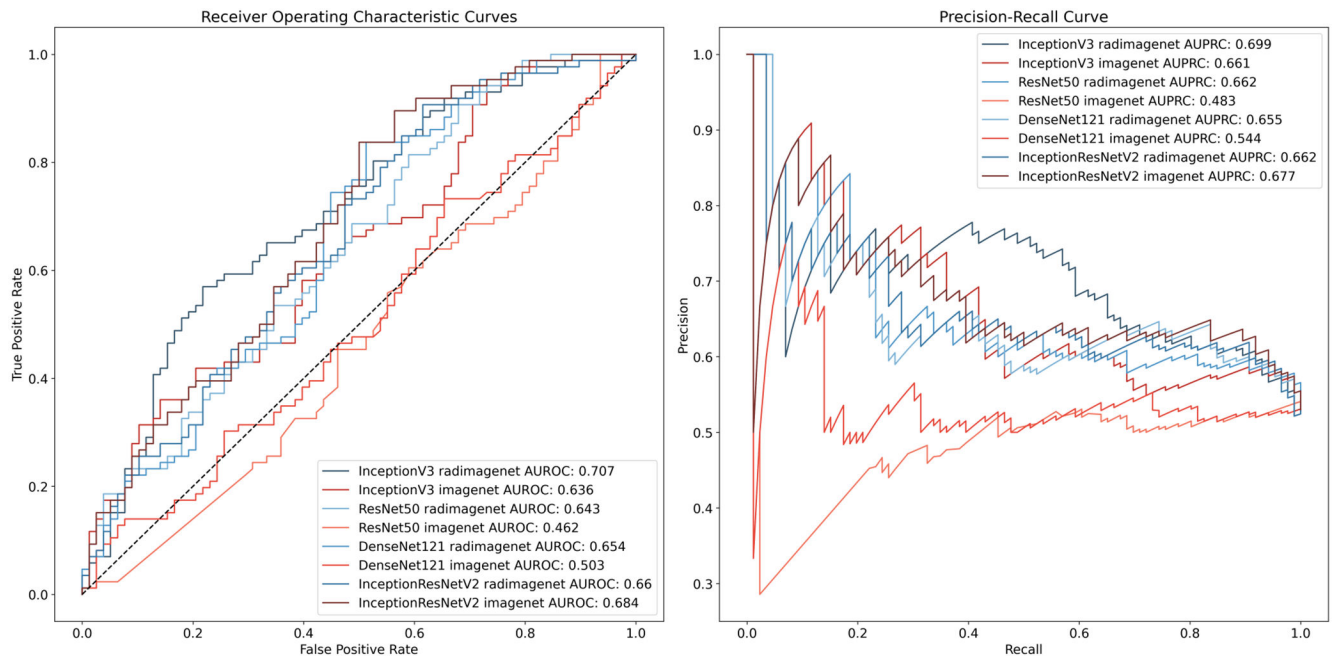


Figure 3. Receiver operating characteristic and precision-recall curves for models predicting margin status

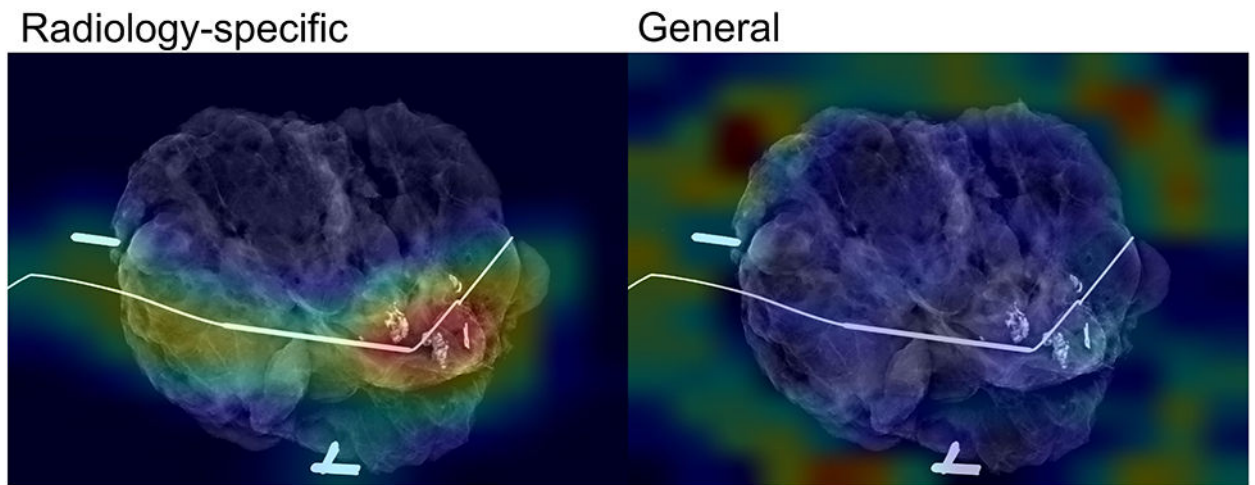


Figure 4. Analysis of pixel importance for radiology-specific and general models. Red/yellow indicates higher importance

Table 1.

Demographic and clinical characteristics

		Overall	Margin (-)	Margin (+)	P-Value	
n		357	139	218		
Age, mean (SD)		60.5 (13.1)	59.2 (12.9)	61.4 (13.2)	0.109	
Race/Ethnicity, n (%)		Asian	11 (3.1)	6 (4.3)	5 (2.3)	0.082
		Hispanic	27 (7.6)	7 (5.0)	20 (9.2)	
		Non-Hispanic Black	63 (17.6)	22 (15.8)	41 (18.8)	
		Non-Hispanic White	253 (70.9)	101 (72.7)	152 (69.7)	
		Other/Unknown	3 (0.8)	3 (2.2)		
Density, n (%)		A	17 (4.8)	6 (4.3)	11 (5.0)	0.651
		B	164 (45.9)	59 (42.4)	105 (48.2)	
		C	130 (36.4)	56 (40.3)	74 (33.9)	
		D	46 (12.9)	18 (12.9)	28 (12.8)	
Tumor Type, n (%)		DCIS	76 (21.3)	30 (21.6)	46 (21.1)	0.925
		IDC	250 (70.0)	96 (69.1)	154 (70.6)	
		ILC	31 (8.7)	13 (9.4)	18 (8.3)	
Grade, n (%)	DCIS, n (%)	1	8 (11.9)	5 (17.9)	3 (7.7)	0.399
		2	31 (46.3)	13 (46.4)	18 (46.2)	
		3	28 (41.8)	10 (35.7)	18 (46.2)	
	IDC, n (%)	1	69 (28.6)	25 (27.2)	44 (29.5)	<0.001
		2	99 (41.1)	26 (28.3)	73 (49.0)	
		3	73 (30.3)	41 (44.6)	32 (21.5)	
	ILC, n (%)	1	7 (22.6)	3 (23.1)	4 (22.2)	0.482
		2	23 (74.2)	9 (69.2)	14 (77.8)	
		3	1 (3.2)	1 (7.7)	0 (0.0)	

BMI – Body mass index, DCIS – Ductal carcinoma in situ, IDC – Invasive ductal carcinoma, ILC – invasive lobular carcinoma

Table 2.

Model accuracy metrics by breast tissue density category

Density Category	AUROC	AUPRC	Sensitivity	Specificity	PPV	NPV
A	0.758	0.877	0.818	0.333	0.692	0.500
B	0.682	0.773	0.875	0.339	0.700	0.606
C	0.756	0.787	0.865	0.357	0.640	0.667
D	0.542	0.614	0.786	0.278	0.629	0.455

Add breast density categories. AUROC – area under the receiver operating characteristic curve, AUPRC – area under the precision-recall curves, PPV – positive predictive value, NPV – negative predictive value

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Model accuracy metrics by race/ethnicity category

Race/Ethnicity Category	AUROC	AUPRC	Sensitivity	Specificity	PPV	NPV
Asian	0.800	0.832	0.800	0.333	0.500	0.667
Hispanic	0.793	0.892	0.900	0.714	0.900	0.714
Non-Hispanic Black	0.737	0.845	0.829	0.364	0.708	0.533
Non-Hispanic White	0.667	0.728	0.861	0.307	0.65	0.596

AUROC – area under the receiver operating characteristic curve, AUPRC – area under the precision-recall curves, PPV – positive predictive value, NPV – negative predictive value