



Multi-magnification-based machine learning as an ancillary tool for the pathologic assessment of shaved margins for breast carcinoma lumpectomy specimens

Timothy M. D'Alfonso¹ · David Joon Ho¹ · Matthew G. Hanna¹ · Anne Grabenstetter¹ · Dig Vijay Kumar Yarlagadda¹ · Luke Geneslaw¹ · Peter Ntiamoah¹ · Thomas J. Fuchs¹ · Lee K. Tan¹

Received: 4 February 2021 / Revised: 11 March 2021 / Accepted: 13 March 2021 / Published online: 26 April 2021

© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2021

Abstract

The surgical margin status of breast lumpectomy specimens for invasive carcinoma and ductal carcinoma in situ (DCIS) guides clinical decisions, as positive margins are associated with higher rates of local recurrence. The “cavity shave” method of margin assessment has the benefits of allowing the surgeon to orient shaved margins intraoperatively and the pathologist to assess one inked margin per specimen. We studied whether a deep convolutional neural network, a deep multi-magnification network (DMMN), could accurately segment carcinoma from benign tissue in whole slide images (WSIs) of shave margin slides, and therefore serve as a potential screening tool to improve the efficiency of microscopic evaluation of these specimens. Applying the pretrained DMMN model, or the initial model, to a validation set of 408 WSIs (348 benign, 60 with carcinoma) achieved an area under the curve (AUC) of 0.941. After additional manual annotations and fine-tuning of the model, the updated model achieved an AUC of 0.968 with sensitivity set at 100% and corresponding specificity of 78%. We applied the initial model and updated model to a testing set of 427 WSIs (374 benign, 53 with carcinoma) which showed AUC values of 0.900 and 0.927, respectively. Using the pixel classification threshold selected from the validation set, the model achieved a sensitivity of 92% and specificity of 78%. The four false-negative classifications resulted from two small foci of DCIS (1 mm, 0.5 mm) and two foci of well-differentiated invasive carcinoma (3 mm, 1.5 mm). This proof-of-principle study demonstrates that a DMMN machine learning model can segment invasive carcinoma and DCIS in surgical margin specimens with high accuracy and has the potential to be used as a screening tool for pathologic assessment of these specimens.

Introduction

Breast-conserving surgery, or lumpectomy, for breast carcinoma, has been shown to be equivalent to mastectomy in terms of survival with the benefit of improved cosmesis [1]. Achieving negative margins (no carcinoma at the margins) is important as positive margins are associated with higher rates of local recurrence [2, 3]. Therefore, accurate microscopic analysis of margins by the pathologist is critical for determining the need for additional surgery and adjuvant radiation.

Various methods are used for orienting and evaluating the margins of lumpectomy specimens [4–6]. The “cavity-

shave” method, the standard technique used at our institution, involves the surgeon excising additional segments of tissue from the wall of the lumpectomy cavity after the main specimen containing the carcinoma has been removed. This method, which allows intraoperative designation of the margin by the surgeon, has been associated with lower rates of positive margins and fewer re-excisions [7–9]. Evaluation of cavity shave margins is the preferred method for many pathologists because it does not require assessment of multiple differentially inked margins of the same specimen. However, microscopic review of these margin specimens can be time-consuming, as it requires examination of multiple additional hematoxylin and eosin (H&E) slides that most often show benign findings.

The development of machine learning approaches, particularly deep learning approaches [10] using convolutional neural networks, along with the rapidly advancing technology of digitalizing histologic slides, could potentially improve the efficiency of the pathologist’s daily workload.

✉ Timothy M. D'Alfonso
dalfonst@mskcc.org

¹ Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

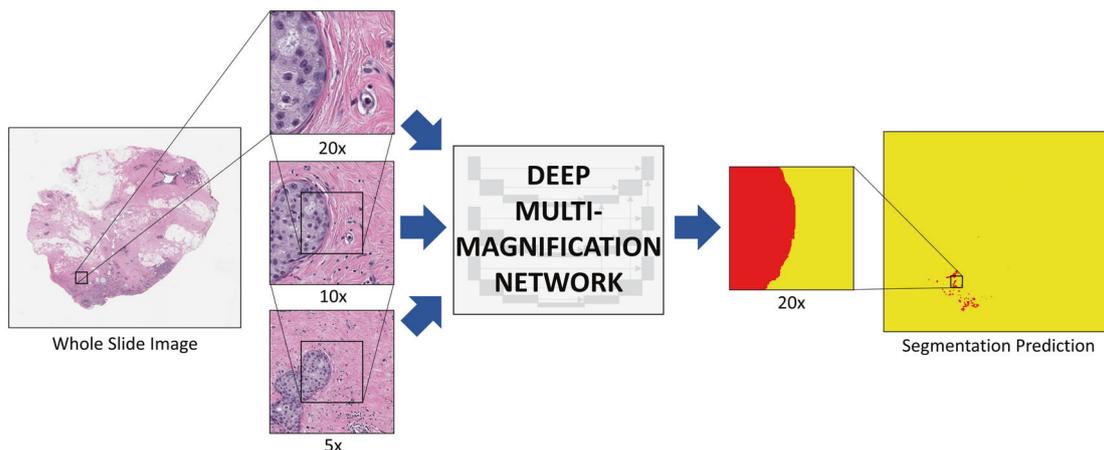


Fig. 1 Deep multi-magnification network (DMMN). Whole slide image (WSI) from a breast margin specimen with DCIS. The DMMN looks at a set of patches from multiple magnifications from the WSI

allowing a wider field-of-view. The segmentation prediction image shows carcinoma highlighted in red, while the remaining tissue is highlighted yellow.

The latter could conceivably be accomplished in a setting in which all histologic slides are scanned prior to review by the pathologist, and in analogy to the automated Papanicolaou screening test, a machine learning model could screen WSIs of margins specimens before manual review, resulting in increased productivity. We used a deep convolutional neural network utilizing morphological features from high magnification and low magnification for more accurate segmentation called deep multi-magnification network (DMMN) [11] to study its potential utility as a screening tool for evaluation of WSIs from breast margin specimens.

Materials and methods

Case selection

After institutional review board approval, lumpectomy specimens containing invasive ductal carcinoma no special type (IDC) and/or DCIS were identified for which all slides in the case had previously been digitally scanned. Invasive lobular carcinoma cases and other special types of breast carcinoma were not studied. Neoadjuvant-treated cases were excluded. We selected cases of IDC and/DCIS of various histologic grades and included cases with IDC or DCIS in the shave margin slides as well as benign cases. All digital WSIs were reviewed by a breast pathologist to confirm the ground truth diagnosis for each WSI.

Digital slide scanning and slide viewer

Glass slides were scanned using Leica Aperio AT2 (Leica Biosystems, Buffalo Grove, Illinois, USA) whole slide scanners at $\times 40$ ($0.25 \mu\text{m}/\text{pixel}$) equivalent magnification using a $20\times/0.75\text{NA}$ Plan Apo objective and $\times 2$ automatic

optical magnification changer. WSIs are accessed through the PICSPlus module in CoPathPlus and launched into the MSK Slide Viewer, an internally developed whole slide scanner, vendor agnostic WSI viewer [12]. The MSK slide viewer allows for standard whole image viewing, slide label viewing, zooming, panning, and slide navigation. The viewer has tools for manual annotation including a ruler for measurements, tracking of viewed slide regions, screenshots, and comments [12].

Tissue segmentation and classification

To segment carcinoma from benign tissue, we used a convolutional neural network called DMMN with multi-encoder, multi-decoder, and multi-concatenation [11] (Fig. 1). The DMMN looks at patches in $20\times$, $10\times$, and $5\times$ magnifications in 256×256 pixels centered at the same coordinate to extract features from both nuclei/texture from high magnification and architectural patterns from low magnification. Our DMMN model can accurately segment carcinomas by utilizing various features from multiple magnifications with a wider field of view. To train our model, we employed a pretrained segmentation model by triple-negative breast cancer (TNBC) dataset as our initial model [11]. We used deep interactive learning (DIAL) [13] to reduce the manual labeling process for finetuning the model. Specifically, we segmented carcinoma in breast margin WSIs using the initial model and manually labeled (or corrected) any false-positive or false-negative regions to update the model. Corrected patches are repeated twice to emphasize them during finetuning.

Once finetuning is complete, the model classifies WSIs as positive or negative for detecting carcinoma. If the number of pixels segmented as carcinoma in a WSI is greater than the set threshold, the image would be classified as positive. The model does not distinguish between invasive carcinoma and

DCIS. All segmentation predictions images were reviewed for confirmation that the segmented pixels represented carcinoma in the WSI. In reviewing segmentation predictions, pixels that are classified as carcinoma are highlighted in red, while the remaining tissue is highlighted in yellow.

Sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve [14] were calculated to assess the performance of the machine learning model in segmenting carcinoma in margin samples.

Results

Cases

The validation set included 408 WSIs (mean, 4.3 WSIs/specimen; range 1–10) from 98 separate shave margin specimens from 20 patients (Table 1). These included 348 (85.3%) benign WSIs and 60 (14.7%) WSIs containing carcinoma. Forty WSIs contained DCIS (6 low-grade, 22 intermediate-grade, 12 high-grade) and 20 WSIs contained IDC (6 well-differentiated, 12 moderately-differentiated, and 2 poorly-differentiated), including 15 WSIs containing both IDC and DCIS.

The testing set, which was independent of the validation set, included 427 WSIs (mean, 4.1 WSIs/specimen; range 1–12) from 104 shave margin specimens from 20 patients (Table 1). These included 374 (87.6%) benign WSIs and 53 (12.4%) WSIs with carcinoma. Forty-one WSIs contained DCIS (five low-grade, 18 intermediate-grade, 18 high-grade) and 12 WSIs contained IDC (three well-differentiated, four moderately-differentiated, five poorly-differentiated), including four WSIs with both IDC and DCIS.

DMMN segmentation predictions

Validation set

The initial model applied to the validation set of WSIs achieved an area under the curve (AUC) = 0.941. A review

of the WSIs that were incorrectly classified as carcinoma revealed that the false positive classifications were due to a variety of reactive changes included biopsy site changes/fat necrosis and stromal elastosis. Other causes of false-positive classification included fibrocystic changes including florid usual ductal hyperplasia, apocrine metaplasia, and cysts/cyst contents. Through further annotation, we corrected those false positive regions on 22 WSIs. After finetuning the model with DiaL, we observed reduced false positives on our segmentation predictions with the AUC = 0.968 (Fig. 2A). We selected the pixel classification threshold to be 10^5 pixels to achieve a sensitivity of 100% (57/57), at which threshold the specificity becomes 78%. With this pixel classification threshold, cancer regions larger than a circle with a radius of 90 μ m would be detected in WSIs.

Testing set

We applied both our initial model and updated model to the testing set of WSIs. Figure 2B shows that the updated model improves the AUC = 0.927 compared to the initial model with the AUC = 0.900. Examples of segmentation predictions for both DCIS and IDC are shown in Figs. 3–5. With the same pixel classification threshold, 10^5 pixels, selected from the validation set, we achieved a sensitivity of 92% (47/51) and specificity of 78% for detecting carcinoma in margin WSIs. We observed four false-negative WSIs in the testing data set. These included a 0.5 mm focus of intermediate-grade DCIS with cautery artifact (Fig. 6A), a 1 mm focus of cribriform low-grade DCIS (Fig. 6B), and two cases in which foci of well-differentiated invasive carcinoma measuring 3 mm and 1.5 mm were present in the margins specimens (Fig. 6C, D, respectively). Of the 374 benign WSIs, 84 (22%) were incorrectly classified as carcinoma. The false positive WSIs resulted from a variety of artifacts such as electrocautery (Fig. 7A, B), tissue folds (Fig. 7C, D), and reactive and benign processes such as biopsy site changes (Fig. 7E, F) and apocrine metaplasia (Fig. 7G, H).

Discussion

We applied a deep convolutional neural network called DMMN to WSIs from breast lumpectomy shave margin specimens and showed that this model could accurately segment carcinoma from benign tissue in these specimens. By using a previously developed segmentation prediction model in a validation set of WSIs from margin specimens followed by correction of incorrect segmentation predictions via manual annotation of additional WSIs, the model achieved high accuracy for detecting carcinoma with an AUC of 0.927, a sensitivity of 92%, and specificity of 78%.

Table 1 Summary of margin diagnoses in whole slide images analyzed in the validation and testing data sets.

Margin diagnosis	Validation data set		Testing data set	
	#WSI	%	#WSI	%
Benign	348	85.3	374	87.6
Ductal carcinoma in situ	40	9.8	41	9.6
Invasive carcinoma	20	4.9	12	2.8
Total	408	100	427	100

WSI whole slide images.

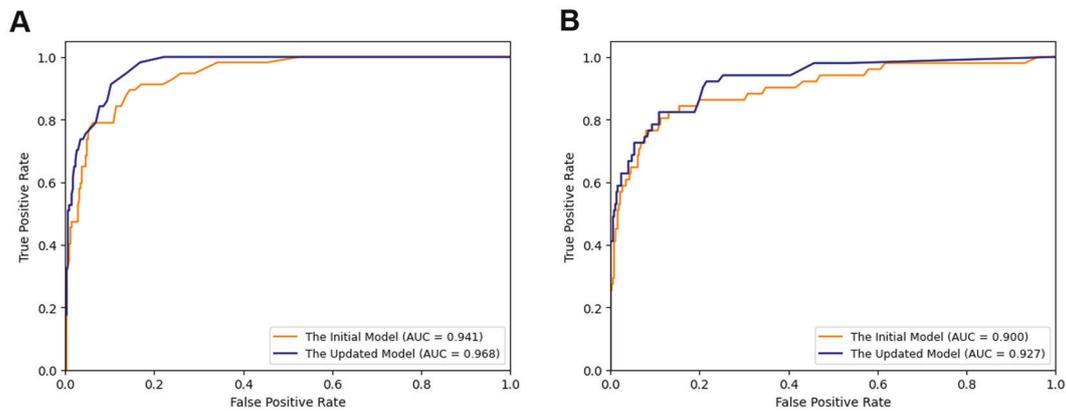
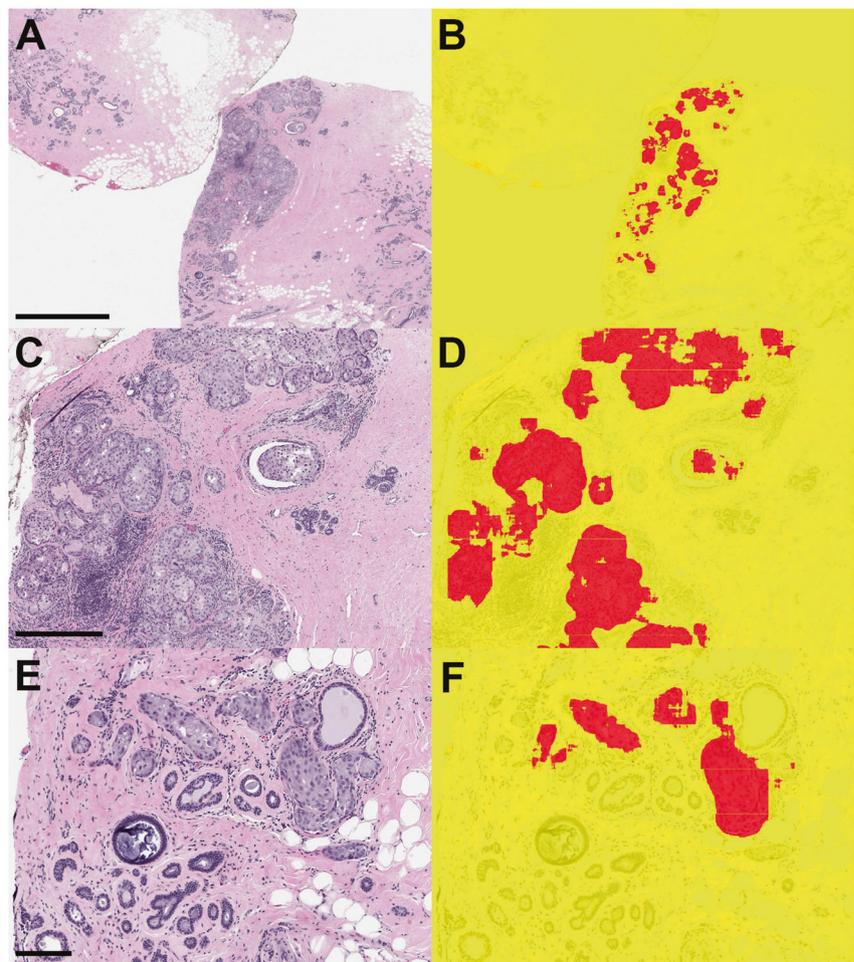


Fig. 2 Receiver-operating characteristic (ROC) curves for the machine learning model's classification of margin specimens. ROC curves for the validation data set (A) and testing data set (B). AUC area under the ROC curve.

Fig. 3 Ductal carcinoma in situ (DCIS) in margin specimen with segmentation classifications. **A** Margin specimen with DCIS (scale bar = 2 mm). **B** Segmentation classification showing DCIS correctly classified (red). Adjacent benign breast glandular tissue is not classified as carcinoma. **C** Higher power view of DCIS (scale bar = 0.5 mm). **D** Segmentation prediction. **E** DCIS involving adenosis (scale bar = 0.2 mm). **F** Segmentation of foci in DCIS in red.



A review of WSIs from the testing data set showed that false-negative classifications resulted from the model classifying 2 small foci of DCIS and 2 foci of well-differentiated invasive carcinoma as benign. Although the performance of the model for identifying carcinoma was excellent in this study, correctly classifying 92% of

carcinoma WSIs, this rate would be unacceptable in actual clinical practice where the model would be used as a screening tool. The false-negative classifications were likely due in part to insufficient training of the model with low-grade carcinomas. Additional training of the model via manual annotation of low-grade DCIS and

Fig. 4 Examples of ductal carcinoma in situ (DCIS) in margin specimens with segmentation classifications.

A DCIS with pagetoid growth in a duct (left) and adjacent benign duct (right) (scale bar = 0.2 mm). **B** Segmentation prediction highlights DCIS in red. **C** High-grade DCIS in margin associated with reactive stroma and chronic inflammation (right) (scale bar = 0.5 mm). **D** Segmentation image showing correct classification of carcinoma.

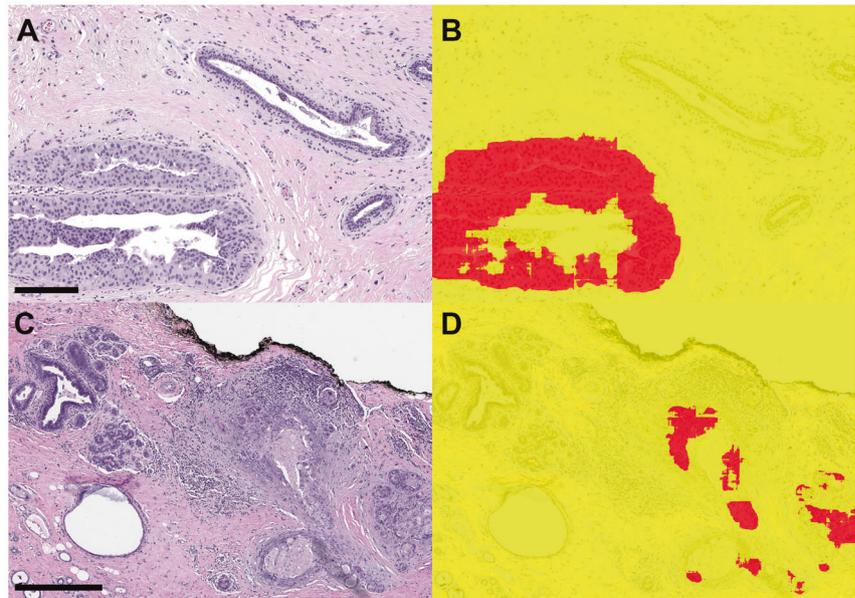
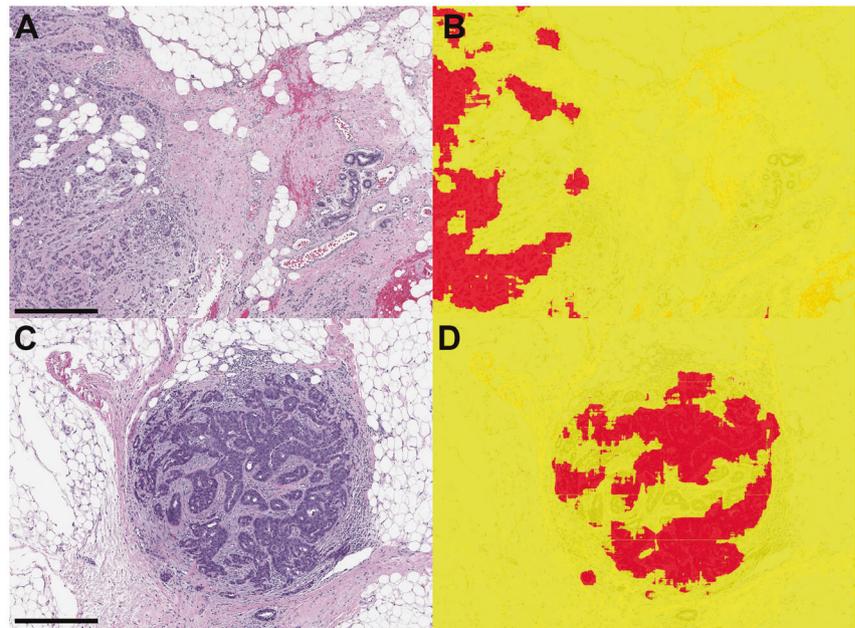


Fig. 5 Invasive ductal carcinoma in margin specimens and segmentation predictions.

A, C Two examples of invasive carcinoma in margin specimens ((**A**): scale bar = 0.5 mm; (**C**): scale bar = 0.5 mm). **B, D** Segmentation predictions show most of the carcinoma was correctly segmented red.

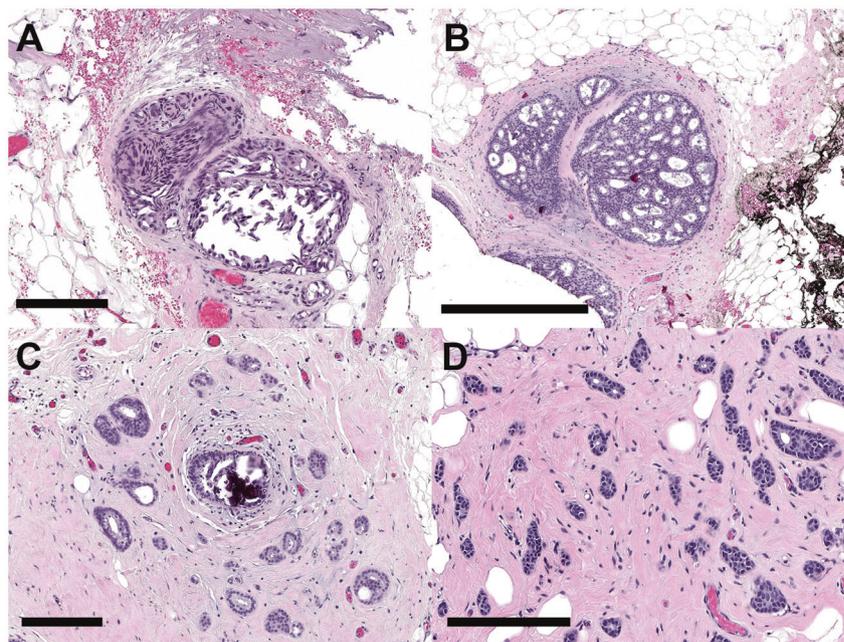


well-differentiated carcinoma is expected to improve its performance.

Convolutional neural networks have gained increasing popularity in computational pathology for carcinoma detection and characterization, and many studies have investigated their utility in breast cancer. The cancer metastases in lymph nodes challenge 2016 (CAMELYON16) was an international competition in which research groups were asked to develop algorithms to detect metastatic carcinoma in sentinel lymph nodes from breast cancer patients [15]. Results from the challenge showed that some convolutional neural network

algorithms achieved a better AUC for identifying lymph node metastases than pathologists when given a time constraint [15]. The follow-up CAMELYON17 challenge showed that convolutional neural network algorithms were also able to stratify metastases according to pN stage, although submitted algorithms were poor in their ability to detect isolated tumor cells [16]. Other studies specific to breast cancer have used convolutional neural networks for quantifying tumor extent [17], mitosis detection [18, 19], predicting grade and molecular subtype [20], and for classifying tumor-associated vs benign stroma in core biopsies [21].

Fig. 6 False negative classifications from the testing set of whole slide images. A Intermediate-grade DCIS with cautery artifact (scale bar = 0.2 mm). **B** Low-grade cribriform DCIS (scale bar = 0.5 mm). **C, D** Two examples of well-differentiated invasive carcinoma ((**C**): scale bar = 0.2 mm; (**D**): scale bar = 0.2 mm).



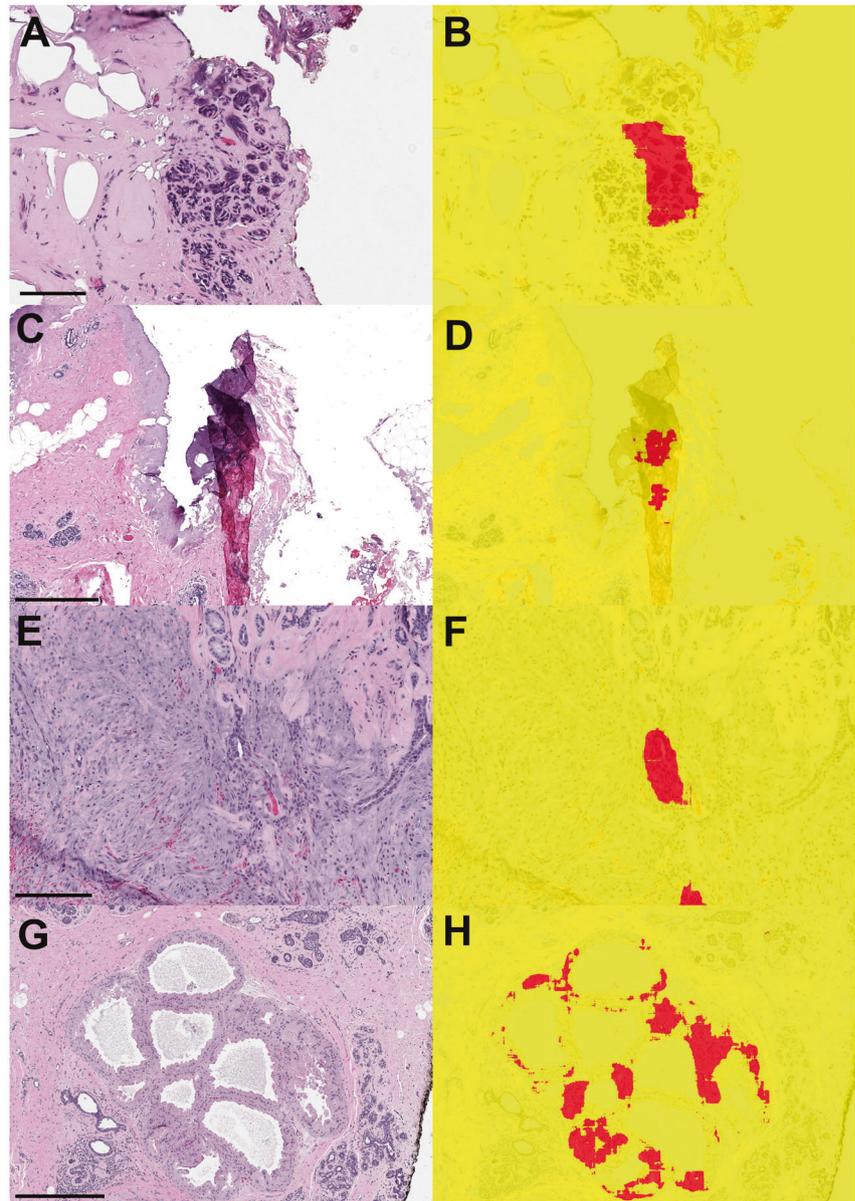
Ours is the first study that specifically used a machine learning model to assess breast shave margin WSIs. The cavity shave method of margin assessment for breast carcinoma lumpectomy specimens is the standard method at our institution and many others. The surgeon removes the entire targeted malignancy in one specimen that is submitted to pathology without orientation. In addition, 5–6 separate margin specimens are submitted that have been taken from the wall of the lumpectomy cavity and are oriented with a stitch indicating the true final margin. The area of the tissue containing the stitch is inked and the specimen is sectioned perpendicular to the true final margin. The pathologist examines multiple histologic sections from these specimens. In a busy breast pathology service, microscopic analysis of these margin specimens can be time-consuming and tedious, and a model such as DMMN could facilitate this process. In such a scenario, the model would be applied to scanned WSIs of margin specimens prior to review by the pathologist. The model could categorize and triage each individual slide as positive or negative for carcinoma and can direct the pathologist to regions containing carcinoma.

To mirror how pathologists examine microscopic slides, we used a convolutional neural network called DMMN in this analysis. This model is unique in that a set of patches from three different magnifications (5 \times , 10 \times , and 20 \times) are used as an input for training the model. Multiple magnifications rather than a patch-based method from one magnification [22] are employed, similar to how a pathologist uses multiple microscopic objectives. To train the model, partial annotation was performed by a pathologist on a set of invasive breast carcinoma WSIs from primary resection

specimens. Tissue types that were annotated include carcinoma (invasive and in situ), benign glandular breast tissue, fibrous stroma, adipose tissue, necrosis, and background (no tissue present). Partial annotation of one WSI takes ~30 min. Following training of the model, multi-class segmentation of breast margin WSIs can be performed using input patches from different magnifications. For this study, we were only interested in segmenting carcinoma from all other tissue types, so we used a binary classification where carcinoma regions were highlighted in red and all non-cancer types are combined and highlighted in yellow.

There are several limitations to our study. We limited our study to cases of IDC and DCIS and did not study invasive lobular carcinomas or other special histologic types of breast carcinoma in this initial study. Invasive lobular carcinomas are typically of low nuclear grade and infiltrate the stroma as single cells. Training of the model was achieved using a set of invasive and in situ ductal carcinomas of various histologic grades that did not include lobular carcinomas. Expansion of the DMMN algorithm to include lobular carcinomas, as well as others, would be necessary to implement this model in practice. This would be true even for cases of IDC as invasive lobular carcinoma may be identified in margin specimens as incidental findings, and thus, the model would need to identify these cases. In addition, our current model does not segment DCIS separately from invasive carcinoma but rather segments carcinoma (DCIS or IDC) from benign tissues. There are also some challenges associated with the evaluation of margin specimens using a machine learning model. Cautery artifact and biopsy site changes are common findings in margins and were sources of false positive classifications in our

Fig. 7 Examples of whole slide images from benign margin specimens that were falsely classified as positive by the machine learning model. **A, B** Benign lobule with cautery artifact ((**A**): scale bar = 0.1 mm). **C, D** Fold in tissue led to incorrect classification with the model ((**C**): scale bar = 0.5 mm). **E, F** Biopsy site with reactive epithelial changes ((**E**): scale bar = 0.2 mm). **G, H** Apocrine metaplasia incorrectly segmented as carcinoma ((**G**): scale bar = 0.5 mm).



analysis. More importantly, however, is that carcinoma in margin specimens usually presents as small foci and it is necessary to have a model sensitive enough to identify minimal carcinoma in these specimens. As we have shown in our validation set, a perfect sensitivity rate is feasible by adjusting the model's pixel threshold. For a model that would be clinically applicable, the ultimate goal must be the elimination of false negatives. Furthermore, future efforts using this DMMN algorithm to quantify carcinoma in the margin as well as measure the distance to the true inked margins will be undertaken.

In summary, we have shown that a machine learning model using patches from multiple magnifications is able to accurately differentiate carcinoma from benign tissue in

WSIs from shave margin specimens of breast lumpectomy specimens with high performance. Using this approach as a screening tool could save significant time for pathologists. The performance of the model in terms of accuracy compared with a pathologist in a true clinical setting, and whether the model improves the efficiency of manual assessment of these specimens requires further study in a prospective setting.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions TMD, DJH, TJF, and LKT conceived the study. DJH, DVKY, LG, and TJF developed the machine learning model. TMD, MGH, AG, and LKT reviewed and annotated WSIs. DJH and PN provided statistical analysis. TMD, DJH, and LKT wrote the first draft. All authors read, edited, and approved the final manuscript.

Funding This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

Compliance with ethical standards

Ethics approval and consent to participate This study was approved by the Institutional Review Board at Memorial Sloan Kettering Cancer Center (Protocol #18-013).

Conflict of interest T.J.F. is co-founder, chief scientist and equity holder of Paige. AI The remaining authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, 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;347:1233–41.
- Houssami N, Macaskill P, Marinovich ML, 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;21:717–30.
- Marinovich ML, Azizi L, Macaskill P, Irwig L, Morrow M, Solin LJ, 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;23:3811–21.
- Molina MA, Snell S, Franceschi D, Jorda M, Gomez C, Moffat FL, et al. Breast specimen orientation. *Ann Surg Oncol.* 2009;16:285–8.
- Wright MJ, Park J, Fey JV, Park A, O'Neill A, Tan LK, et al. Perpendicular inked versus tangential shaved margins in breast-conserving surgery: does the method matter? *J Am Coll Surg.* 2007;204:541–9.
- Guidi AJ, Connolly JL, Harris JR, Schnitt SJ. The relationship between shaved margin and inked margin status in breast excision specimens. *Cancer.* 1997;79:1568–73.
- Cao D, Lin C, Woo SH, Vang R, Tsangaris TN, Argani P. Separate cavity margin sampling at the time of initial breast lumpectomy significantly reduces the need for reexcisions. *Am J Surg Pathol.* 2005;29:1625–32.
- Chagpar AB, Killelea BK, Tsangaris TN, Butler M, Stavris K, Li F, et al. A randomized, controlled trial of cavity shave margins in breast cancer. *N Engl J Med.* 2015;373:503–10.
- Moo TA, Choi L, Culpepper C, Olcese C, Heerdt A, Sclafani L, et al. Impact of margin assessment method on positive margin rate and total volume excised. *Ann Surg Oncol.* 2014;21:86–92.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–44.
- Ho DJ, Yarlagaadda DVK, D'Alfonso TM, Hanna MG, Grabenstetter A, Ntiamoah P, et al. Deep multi-magnification networks for multi-class breast cancer image segmentation. *Comput Med Imaging Graph.* 2021;88:101866.
- Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol.* 2019;32:916–28.
- Ho DJ, Agaram NP, Schuffler PJ, Vanderbilt CM, Jean MH, Hameed MR, et al. Deep interactive learning: an efficient labeling approach for deep learning-based osteosarcoma treatment response assessment. Springer; 2020.
- Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* 2018;63:07TR01.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* 2017;318:2199–210.
- Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermsen M, et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Trans Med Imaging.* 2019;38:550–60.
- Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. *Sci Rep.* 2017;7:46450.
- Wang H, Cruz-Roa A, Basavanahally A, Gilmore H, Shih N, Feldman M, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *J Med Imaging.* 2014;1:034003.
- Malon CD, Cosatto E. Classification of mitotic figures with convolutional neural networks and seeded blob features. *J Pathol Inf.* 2013;4:9.
- Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer.* 2018;4:30.
- Bejnordi BE, Lin J, Glass B, Mullooly M, Gierach GL, Sherman ME, et al. Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images. *Proc IEEE Int Symp Biomed Imaging.* 2017:929–32.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Springer; 2015.