Breast Cancer Research and Treatment (2021) 186:379–389
https://doi.org/10.1007/s10549-020-06093-4

PRECLINICAL STUDY

Analysis of tumor nuclear features using artificial intelligence to predict response to neoadjuvant chemotherapy in high-risk breast cancer patients

David W. Dodington[1] · Andrew Lagree[2] · Sami Tabbarah[3] · Majid Mohebpour[2] · Ali Sadeghi-Naini[3,4] · William T. Tran[2,3,5,7] · Fang-I Lu[1,6]

Received: 22 October 2020 / Accepted: 31 December 2020 / Published online: 23 January 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Purpose Neoadjuvant chemotherapy (NAC) is used to treat patients with high-risk breast cancer. The tumor response to NAC can be classified as either a pathological partial response (pPR) or pathological complete response (pCR), defined as complete eradication of invasive tumor cells, with a pCR conferring a significantly lower risk of recurrence. Predicting the response to NAC, however, remains a significant clinical challenge. The objective of this study was to determine if analysis of nuclear features on core biopsies using artificial intelligence (AI) can predict response to NAC.

Methods Fifty-eight HER2-positive or triple-negative breast cancer patients were included in this study (pCR $n=37$, pPR $n=21$). Multiple deep convolutional neural networks were developed to automate tumor detection and nuclear segmentation. Nuclear count, area, and circularity, as well as image-based first- and second-order features including mean pixel intensity and correlation of the gray-level co-occurrence matrix (GLCM-COR) were determined.

Results In univariate analysis, the pCR group had fewer multifocal/multicentric tumors, higher nuclear intensity, and lower GLCM-COR compared to the pPR group. In multivariate binary logistic regression, tumor multifocality/multicentricity (OR = 0.14, $p=0.012$), nuclear intensity (OR = 1.23, $p=0.018$), and GLCM-COR (OR = 0.96, $p=0.043$) were each independently associated with likelihood of achieving a pCR, and the model was able to successful classify 79% of cases (62% for pPR and 89% for pCR).

Conclusion Analysis of tumor nuclear features using digital pathology/AI can significantly improve models to predict pathological response to NAC.

Keywords Breast cancer · Neoadjuvant chemotherapy · Artificial intelligence · Digital pathology

✉ Fang-I Lu
fangi.lu@sunnybrook.ca

[1] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

[2] Biological Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada

[3] Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[4] Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada

[5] Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada

[6] Department of Laboratory Medicine and Molecular Diagnostics, Sunnybrook Health Sciences Centre, 2075 Bayview Ave., Rm E423a, Toronto, ON M4N 3M5, Canada

[7] Temerty Centre for Artificial Intelligence Research and Education in Medicine, University of Toronto, Toronto, ON, Canada

Introduction

Neoadjuvant chemotherapy (NAC) has become an integral part of the multidisciplinary approach to breast cancer treatment [1, 2]. One of the benefits of NAC is that it allows for the definitive evaluation of treatment response post-operatively. The treatment response can be classified as either a pathological partial response (pPR) or a pathological complete response (pCR) defined as no residual invasive or in situ disease in the breast or lymph nodes [3]. Multiple studies have shown that achieving a pCR after NAC is associated with improved event-free survival and overall survival, particularly for triple-negative and HER2-positive tumors [3–5]. Unfortunately, pCR rates are highly variable and unpredictable, and there is great interest in developing tools to predict response to NAC a priori in order to identify patients that might benefit from novel treatment approaches in the neoadjuvant setting.

Several studies have sought to identify predictive signatures using clinical, radiologic, pathologic, and molecular markers. HER2-positive and triple-negative tumors have consistently demonstrated increased likelihood of a pCR [5–9]. Higher histologic grade has been associated with better response to NAC in several studies [7, 8, 10] but not all [11]. Tumor-infiltrating lymphocytes (TILs) have also been reported by other investigators as predictors of pCR [10–12] as well as a higher Ki67 proliferation rates [6, 7]. Despite these promising findings, other studies, which have examined over twenty histological and immunohistochemical markers, have failed to find any significant predictors of response to NAC in triple-negative and HER2-positive tumors [13, 14].

Artificial intelligence (AI) and its subfields of machine learning and deep learning are quickly becoming powerful tools in medicine and oncology. These technologies have shown promise in the field of breast cancer for both diagnostic and prognostic purposes [15–17]. The objective of this research was to develop a quantitative digital pathology imaging (QDPI) platform that employs AI to map tumor components and extract predictive biomarkers from breast core needle biopsies. We hypothesized that automated tumor bed mapping, nuclei segmentation, and quantification of tumor nuclear features would permit development of predictive models of response to NAC.

Materials and methods

Study population

This study was approved by the institutional research ethics board. A retrospective cohort was established consisting of breast cancer patients treated with NAC between 2007 and 2018 at Sunnybrook Health Sciences Centre (Toronto, Canada) (Fig. 1). All patients had a breast core needle biopsy prior to NAC with pathological review at our institution. The initial cohort consisted of a total of 74 cases (36 pPR and 38 pCR cases). Among these, 73 of 74 were diagnosed as invasive mammary carcinoma or invasive ductal carcinoma

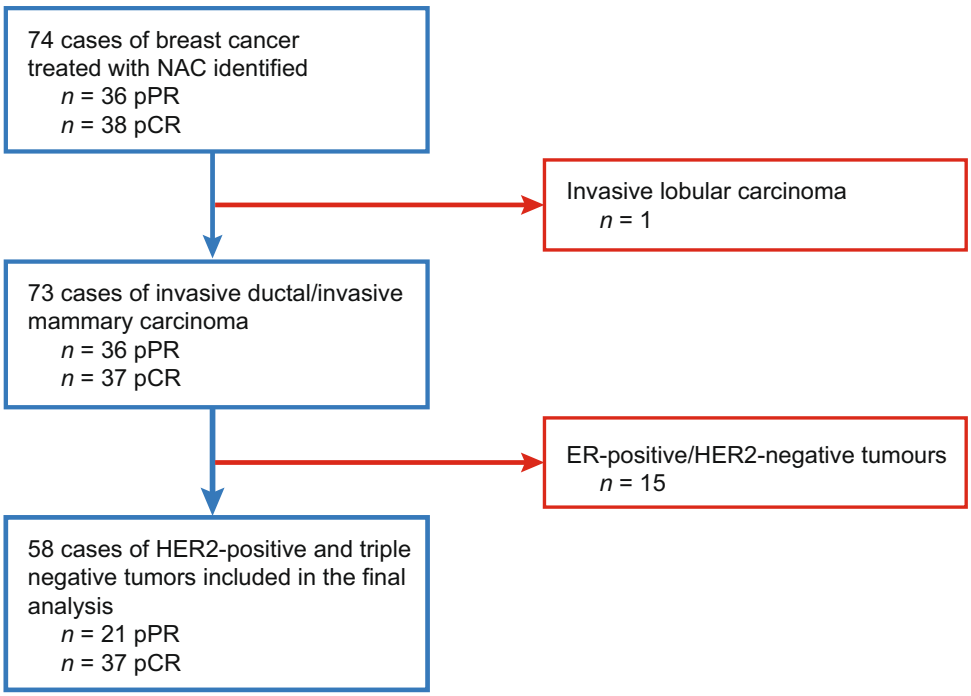

Fig. 1 Flowchart of cases. *NAC* neoadjuvant chemotherapy, *pPR* pathologic partial response, *pCR* pathologic complete response

not otherwise specified. There was a single case of invasive lobular carcinoma, which was not included in further analysis. For hormone receptor status, ER negative and ER low positive (staining in 1–10% of tumor cells) were grouped together as ER-negative, since ER-low positive tumors are more similar to ER-negative ones in terms of biology and response to NAC [18, 19]. HER2 status was determined by immunohistochemistry followed by fluorescence in situ hybridization in cases with equivocal staining. The initial cohort included ER-positive/HER2-negative tumors ($n=15$), HER2-positive tumors with or without ER-positivity ($n=38$) and triple-negative tumors ($n=20$). All ER-positive/HER2-negative tumors had a pPR and were excluded in our analysis. All patients received anthracycline-taxane-based NAC. Pre-treatment tumor size was determined based on the maximum dimension reported from MRI or ultrasound. Response to NAC was determined on the final surgical excision specimen and was defined using a binary classification; a pCR was defined as no residual invasive carcinoma present in the breast or lymph nodes after presurgical therapy [20]. All clinical and pathological data were collected from the electronic patient record.

Core biopsy sample preparation

Core biopsy specimens were sectioned into 4 µm microtomes from formalin-fixed paraffin-embedded (FFPE) blocks. Specimens were prepared onto glass slides and stained with hematoxylin & eosin (H&E). The slides were digitized into whole slide images (WSI) using a TissueScope LE digital pathology image scanner (Huron Digital Pathology Inc., St. Jacobs, Canada) at 40× magnification. All WSIs were reviewed to ensure image integrity before image processing and analysis.

Image processing algorithm

Tumor region classification in digitized core biopsy samples

Tumor bed regions of the digitized core biopsies were identified using a deep learning-based classification approach. A training dataset was prepared with ground-truth labels. A board-certified breast pathologist analyzed the digitized pathology images. Each WSI was manually reviewed, and the tumor bed regions of interest (ROI) were annotated using Sedeen Viewer (Pathcore Inc., Toronto, Canada). The ROIs were delineated based on an area that demonstrated at least 50% tumor cellularity. The ROIs were isolated from the WSI, then tiled to a size of 750×750 pixels with no overlap. The tiled images were color normalized [21], to correct color variations and intensity differences between slides.

A deep convolutional neural network (CNN) was developed to automate tumor bed detection of WSI (Fig. 2). The network was adapted from Visual Geometry Group (VGG) 19 [22] architecture with weights pre-trained on the ImageNet dataset [23]. The network took inputs of 750×750×3 pixel H&E images and passed them through repeated blocks of 3×3 convolutions with a rectified linear unit (ReLU) activation [24] and max-pooling layers. The network's final layers featured a global average pooling layer [25] and a dense layer with SoftMax [26] activation. The network contained 17 weighted layers in total.

Thirty WSI were randomly selected to train and test the classification network. The WSI were randomly split at a

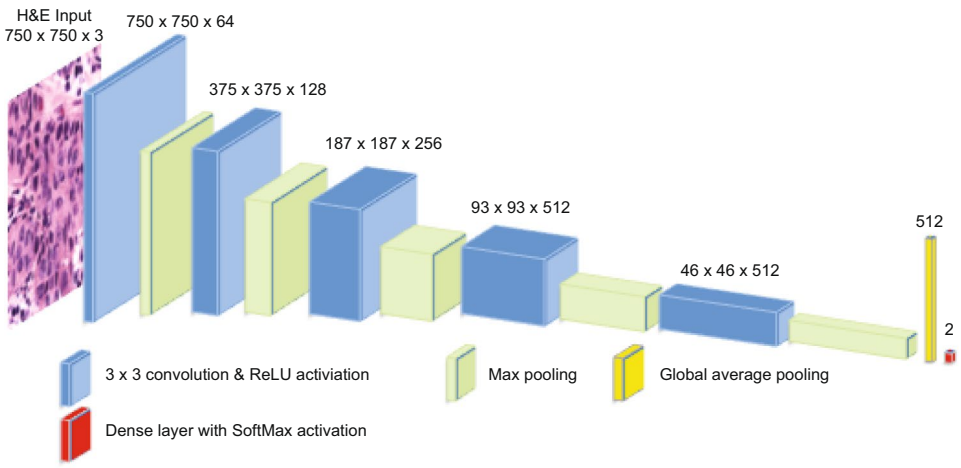

Fig. 2 Deep convolutional neural network architecture. A deep CNN was developed to automate tumor bed detection of WSI. The network was adapted from VGG 19 and used weights pre-trained on the ImageNet dataset. The network took inputs of 750×750×3 pixel H&E images and passed them through repeated blocks of 3×3 convolutions with a ReLU activation and max-pooling layers. The network's final layers featured a global average pooling layer and a dense layer with SoftMax activation. The network contained 17 weighted layers in total

ratio of 80:20, with approximately 6200 images in the training set and 1500 images in the test set. A board-certified breast pathologist reviewed all images selected for the datasets. We provided the network with images from two classes: (1) tumor bed versus (2) benign tissue. Images of the tumor bed consisted of previously extracted tiles. Benign tissue images were extracted from regions outside of the annotated ROI, contained no more than 10% white background, and were color normalized [21].

Tumor nuclei segmentation

Three deep CNNs, based on the U-Net [27] architecture, were trained, for segmentation of tumor cell nuclei. U-Net has shown to provide excellent biomedical semantic segmentation results in previous studies [28]. We further applied custom weighted loss functions that heavily penalized incorrect cell nuclei boundaries to improve correct classification of adjacent or overlapping nuclei. The standard CNN encoders of networks one and two were replaced with residual networks [29] (ResNet) 152, whereas network three implemented ResNet 101. Furthermore, networks one and two were initialized with weights pre-trained on the ImageNet dataset [23] and network three was trained from scratch. The decoder, within each network, consisted of up-sampling, then concatenation of the respective feature map from the encoder, followed by two blocks of 3×3 convolutions, batch normalization, and ReLU activation [24]; the final layer featured a 1×1 convolution with SoftMax activation [26]. Weighted cross-entropy loss [27] was implemented in network one, while networks two and three had a combined loss function of weighted cross-entropy [27] and soft dice. To improve prediction accuracy, we ensembled the probability maps of the networks using a pixel-wise mean.

Open-sources images from the Multi-Organ Nucleus Segmentation challenge [30–32] were utilized to create training and test sets for the networks. The dataset of the challenge was curated from the cancer genome atlas (TCGA) database and contained 44 H&E stained, 1000×1000 pixel images from nine organs, imaged at 40× magnification [32], of which approximately 28,000 nuclei boundaries of epithelial and stromal cells were manually annotated and approved by an expert pathologist. The images from the database were randomly split at a ratio of 80:20 between training and test sets, respectively. As we implemented weighted loss functions, the weighted map of each training and validation image was calculated. Subsequently, we implemented image augmentation [33] to improve the network's generalizability and performance.

Feature extraction and selection

Using a software toolkit (HistomicsTK) [34], 48 features were extracted, per image tile for analyses. The features were representative of four categories including nuclear size, nuclear intensity, nuclear shape, and Haralick parameters [35]. The predicted binary masks (Fig. 3e) and respective gray-scale image tiles (Fig. 3d) were used to calculate nuclear intensity and Haralick parameters. Among these features, five were handpicked in order to represent the following histological features: cellularity (nuclear count), nuclear enlargement (nuclear size area), nuclear irregularity (nuclear circularity), hyperchromasia (mean nuclear intensity), and nuclear texture (gray-level co-occurrence matrix correlation). For these five features, we performed univariate analyses comparing pCR and pPR (independent *t* test, two sided, $p < 0.10$), and significant variables were retained for predictive modeling.

Statistical analysis

Statistical analysis was performed using SPSS version 20 (IBM Corp., Armonk, NY) and Prism GraphPad version 6 (GraphPad Software, San Diego, CA). Univariate comparisons were performed using a two-tailed independent Student's *t* test for continuous variables and chi-square test for categorical variables. Multivariate analysis was performed using logistic regression for pCR vs. pPR binary outcomes. Clinicopathological and nuclear features that were significantly different in univariate analysis ($p < 0.10$) were retained for multivariate modeling. Odds ratios are presented for each predictor. Statistical significance was defined as $p < 0.05$.

Results

Patients and clinicopathological variables

The analyzed cohort consisted of 21 patients with a pPR and 37 patients with a pCR. Clinical and pathological data by group are shown in Table 1. The only significant difference between the two groups was that the pCR group had fewer patients with multifocal and/or multicentric disease. The association between multifocal/multicentric disease and decreases rates of pCR have been previously documented [36].

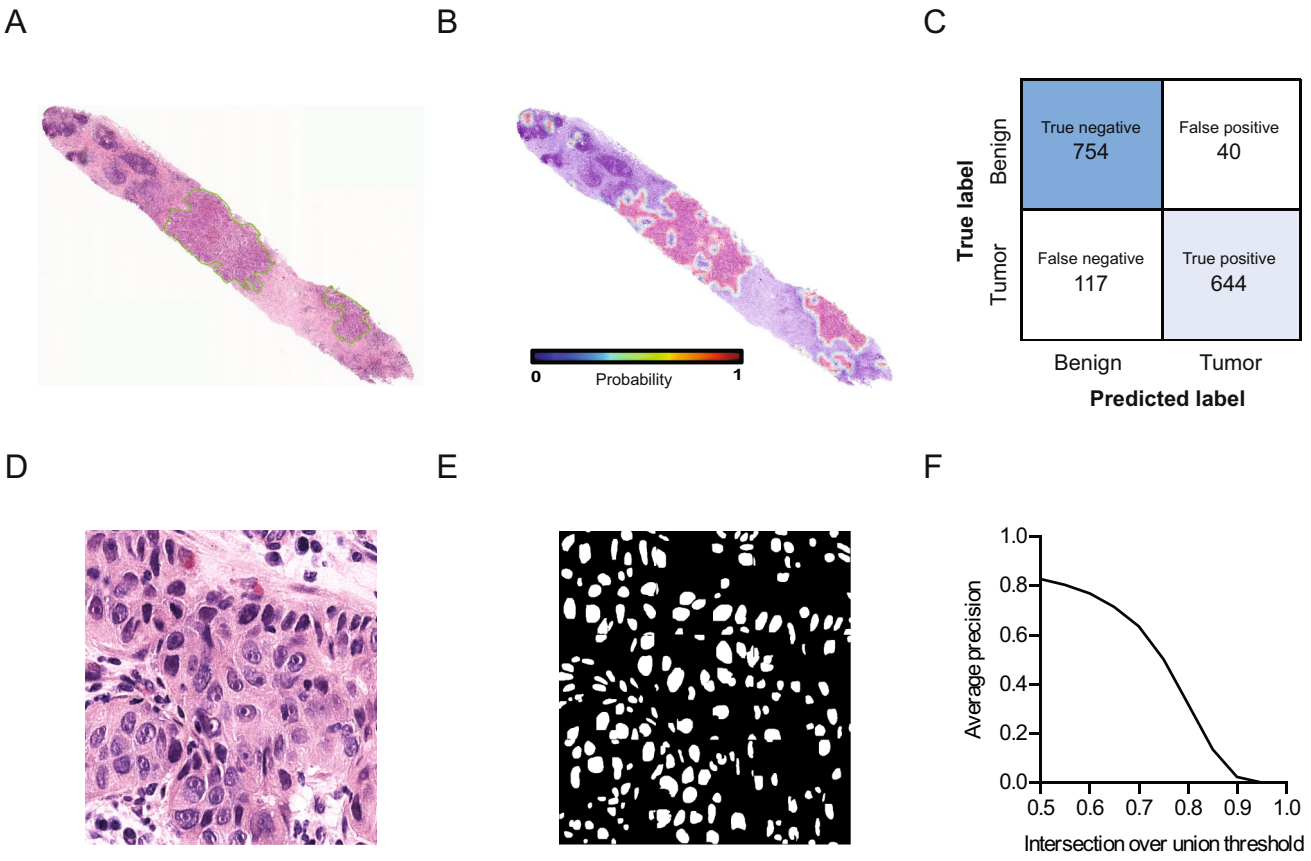

Fig. 3 Automatic tumor bed detection and nuclear segmentation using CNNs. (**a**) Representative ground-truth annotation of breast core needle biopsy, outlining tumor bed ROI. (**b**) Representative WSI heatmap displaying tumor bed probabilities. (**c**) Confusion matrix showing accuracy of CNN for predicting tiles containing tumor versus benign tissue in the test set (**d**) Representative image tile extracted from tumor bed, 750 x 750 pixels, 400× magnification. (**e**) Predicted binary mask of cell nuclei locations from (**d**). (**f**) Average precision versus IoU threshold for nuclear segmentation

Automatic detection of tumor bed and nuclear segmentation using CNNs

Representative tumor bed annotations of a core needle biopsy and heatmap displaying automated tumor bed probabilities are shown in Fig. 3a and b, respectively. Overall, the modified VGG network was 90% accurate in classifying tumor versus benign tissue (Fig. 3c). For the tumor class, the precision was 0.94, recall was 0.85, and F1-score was 0.89. For the benign tissue class, the precision was 0.87, recall was 0.95, and F1-score was 0.91. Overall the F1-score was 0.89, and F1-beta was 0.92.

Representative H&E stained tile and corresponding binary mask of cell nuclei locations are shown in Fig. 3d and e, respectively. In the test set, the ensembled results displayed 93% pixel-wise accuracy. Mean average precision of the ensembled results was calculated to be 0.47, across intersection over union (IoU) thresholds beginning at 0.5 and increasing linearly by 0.05–0.95 (Fig. 3f). At an IoU, threshold of 0.7 precision was 0.63, recall was 0.57, and F1 was 0.60.

Feature reduction

Next, we compared the extracted nuclear features in tumors from patients with partial and complete pathological responses. We examined five features including nuclear count, area, circularity, intensity, and GLCM-COR, a Haralick-based texture feature that is a measure of how similar a pixel's intensity is to its neighbors. In univariate analysis, tumors from patients with a pCR had a significantly higher nuclear intensity (Fig. 4a) and lower GLCM-COR (Fig. 4b) compared to patients with a pPR. There was no significant difference in nuclear count (Fig. 4c), nuclear area (Fig. 4d), or circularity (Fig. 4e); therefore, these features were not used for predictive modeling.

Table 1 Clinicopathological characteristics of the study population[a]

	pPR ($n=21$)	pCR ($n=37$)	p Value[b]
Diagnosis			
Invasive ductal carcinoma	19 (91%)	37 (100%)	0.13
Invasive mammary carcinoma	2 (9%)	0 (0%)	0.13
Age (years)	50 ± 9 (36–67)	50 ± 10 (30–72)	0.97
Postmenopausal	12 (57%)	17 (46%)	0.82
Tumor size (cm)	9.68 ± 17.80 (2.00–73.00)	9.59 ± 12.78 (1.25–45.00)	0.98
Tumor extent			
Locally advanced	20 (95%)	36 (97%)	1.00
Early breast cancer	1 (5%)	1 (3%)	1.00
Multifocal and/or multicentric disease	11 (52%)	9 (24%)	**0.031**
Skin involvement	2 (10%)	5 (14%)	0.71
Inflammatory breast cancer	1 (5%)	4 (11%)	0.64
Positive lymph node(s)	16 (76%)	26 (70%)	0.63
Treatment regime			
AC-T chemotherapy	9 (43%)	24 (65%)	0.10
FEC-D chemotherapy	12 (57%)	13 (35%)	0.10
Cycles (number)	6 ± 2 (3–10)	7 ± 2 (3–11)	0.50
Receptor status			
ER positive/HER2 positive	11 (52%)	14 (38%)	0.28
ER negative/HER2 positive	3 (14%)	10 (27%)	0.34
ER negative/HER2 negative	7 (33%)	13 (35%)	0.89
Tumor grade			
Grade 1	0 (0%)	1 (3%)	–
Grade 2	9 (43%)	11 (30%)	0.43
Grade 3	12 (57%)	25 (68%)	0.43

Bold value indicates the statistical significance of $p < 0.05$

ER estrogen receptor, *PR* progesterone receptor, *AC-T* adriamycin (doxorubicin) cyclophosphamide taxol, *FEC-D* fluorouracil epirubicin cyclophosphamide docetaxel

[a]Data represent mean ± standard deviation (range) for continuous variables and count (%) for categorical variables

[b]Associations were evaluated using independent Student's *t* test for continuous variables and chi-square test for categorical variables

Predictive modeling of response to NAC

For multivariate analysis, a baseline logistic regression model was created to predict the likelihood of a pCR using clinicopathological variables first. Given that disease focality was the only significant variable in univariate analysis, the baseline logistic regression model included only this variable (Table 2). The baseline model was statistically significant, and the presence of multifocal/multicentric disease was associated with decreased odds of achieving a pCR. The model was able to correctly classify 67% of cases (52% for pPR and 76% for pCR).

Next, in a two-step logistic regression, both nuclear intensity and GLCM-COR were each tested as predictors of a pCR in addition to disease focality. In *model 1*, addition of nuclear intensity to tumor focality provided a significant increase in the chi-square value, and higher nuclear intensity was associated with increased odds of a pCR. This model was able to correctly classify 71% of cases (52% for pPR and 81% for pCR). In *model 2*, addition of GLCM-COR to tumor focality provided a significant increase in the chi-square value, and higher GLCM-COR was associated with decreased odds of a pCR. This model was able to correctly classify 76% of cases (52% for pPR and 89% for pCR). In *model 3*, adding GLCM-COR to a model already containing tumor focality and nuclear intensity resulted in a significant increase in the Chi-square value. This model was able to correctly classify 79% of cases (62% for pPR and 89% for pCR). Notably, this model was the only one that improved classification (data not shown accuracy for pPR cases). To test for biased prediction due to class imbalance, we separately performed both random under-sampling of the over-represented group and random up-sampling of the under-represented group. This analysis did not alter the statistical significance of the predictors in the model or the overall classification accuracy (data not shown).

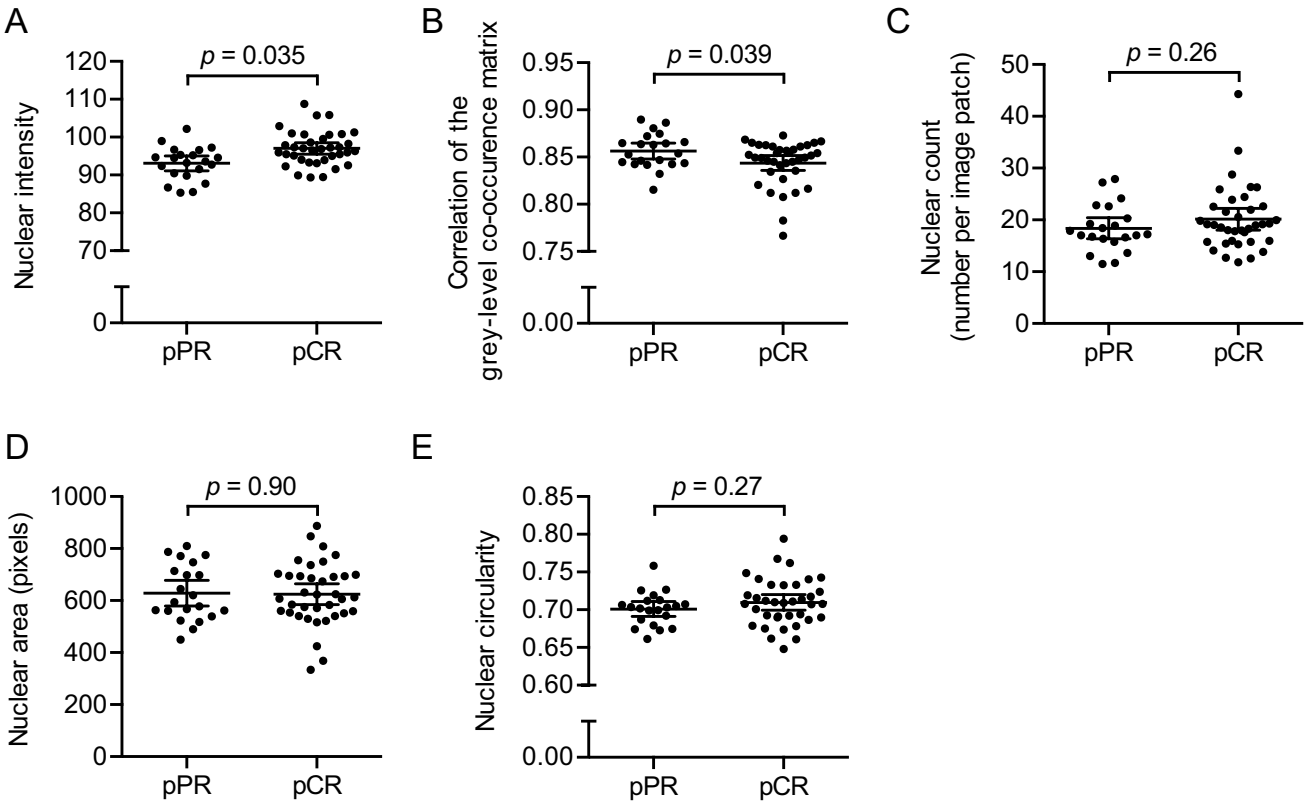

Fig. 4 Comparison of tumor nuclear features prior to NAC in patients achieving partial and complete pathological responses. (**a**) Nuclear intensity, (**b**) correlation of the gray-level co-occurrence matrix (GLCM-COR), (**c**) nuclear count, (**d**) nuclear size, and (**e**) nuclear circularity in tumors from patients achieving a pathological partial response (pPR) and pathological complete response (pCR). Data represent mean ± 95% confidence interval. Groups were compared using two-tailed Student's *t* test

Discussion

This study is the first to predict the response of breast cancer to NAC using well-established machine learning tools. Multiple CNNs were able to automate tumor bed detection and nuclear segmentation. We evaluated five pre-determined features and found among them, nuclear intensity and a texture-based features, GLCM-COR, were able to significantly predict the likelihood of a patient achieving a pCR. Importantly, these markers provided novel predictive information that was not accounted for in standard clinicopathological variables, thus, providing a strong rationale for including AI in routine evaluation of core needle biopsies.

Predicting tumor behavior and prognosis is one of the main objectives in breast pathology, which has the potential to be greatly enhanced with the use of AI and machine learning [16]. One particular area where AI has been investigated is for the histologic grading of tumors. Conventionally, pathologists use the Nottingham score to grade breast tumors, which involves analysis of three components: glandular differentiation, nuclear pleomorphism, and mitotic rate. Work has been done to address all three of these aspects. Romo-Bucheli et al. used a deep learning classifier to automatically identify tubule nuclei as a means of standardizing quantification of glandular differentiation [37]. They found that the ratio of tubular nuclei to total nuclei significantly predicted Oncotype Dx score, which predicts the chances of recurrence in ER-positive breast cancer. Several studies have also used machine learning to aid pathologists in counting mitotic figures. For example, in the International Conference on Pattern Recognition 2012 mitosis detection competition, two groups were able to identify mitotic figures in breast samples with high accuracy; Ciresan et al. [38] used deep max-pooling CNN with an *F*-measure of 0.78, and Wang et al. [39] used a combined CNN and handcrafted features (morphology, color, and texture) with an *F*-measure of 0.73. At last, AI can be particularly useful for quantification of nuclear pleomorphism, one of the more subjective components of Nottingham grading. Veta at al. employed an automatic nuclei segmentation algorithm to evaluate nuclear area and found this to be inversely associated with patient survival [40]. Whitney et al. extracted 216 nuclear features from both epithelial and stromal cells and used feature ranking to predict the Oncotype DX risk score where they found an accuracy of 75–86% in discriminating low- vs. high-risk score [41]. The features identified in our study are perhaps

Table 2 Two-step binary logistic regression estimating likelihood of pCR to NAC

	Baseline Step 1: Multifocal/multicentric		*Model 1* Step 1: Multifocal/multicentric Step 2: Intensity		*Model 2* Step 1: Multifocal/multicentric Step 2: GLCM-COR		*Model 3* Step 1: Multifocal/multicentric, intensity Step 2: GLCM-COR	
Model statistics								
Nagelkerke R^2	0.11		0.32		0.27		0.41	
Chi-square	4.61		15.28		12.89		20.68	
p value	0.032		<0.001		0.002		<0.001	
Chi-square change	–		10.67		8.28		5.40	
p value change	–		0.001		0.004		0.020	
Correct classification								
pPR	52.4%		52.4%		52.4%		61.9%	
pCR	75.7%		81.1%		89.2%		89.2%	
Overall	67.2%		70.7%		75.9%		79.3%	
Predictors	OR (95% CI)	*p* value	OR (95% CI)	*p* value	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Multifocal and/or multicentric	0.29 (0.09–0.91)	0.034	0.27 (0.08–0.95)	0.042	0.16 (0.04–0.63)	0.009	0.14 (0.03–0.65)	0.012
Intensity			1.24 (1.06–1.46)	0.008			1.23 (1.04–1.47)	0.018
GLCM-COR					0.96 (0.92–0.99)	0.016	0.96 (0.92–0.99)	0.043

GLCM-COR correlation of the gray-level co-occurrence matrix, *OR* odds ratio

implicitly related to histologic grade. Greater variability in nuclear texture (low GLCM-COR measure) is likely associated with the lack of uniformity of chromatin commonly seen in higher grade tumors, although the color/intensity of nuclei is not a specific feature used to determine the nuclear pleomorphism score in Nottingham grading. Interestingly though, in our cohort, tumor grade as evaluated by pathologists was not predictive of response to NAC. A higher histologic grade has been a significant predictor of a pCR in some studies [7, 8, 10] but not all [11]. In our case, this may be due to a relatively small sample size, and we were not able to detect the small to moderate effect size of grade. Alternatively, it may be that grading, which reduces a set of complex continuous features into a small number of categories results in loss of important information. This is supported by the fact that analysis of only 1 or 2 nuclear features in our relatively small cohort outperformed traditional grading for predicting response to NAC. This speaks to the potential advantages of using machine learning tools as it allows for reproducible quantification of these complex and sometimes sub-visual features, which can provide important prognostic information.

The two features found to be associated with response to NAC in our study were nuclear intensity and GLCM-COR, which are reportedly the most common and sensitive texture descriptor [42]. Lu et al. evaluated 615 features relating to nuclear shape/texture and found that the intensity range, rather than the mean intensity value, was among the top 15 features capable of predicting poorer survival [43]. Chen et al. analyzed 730 morphological features including nuclear intensity and Haralick-based texture features; however, neither of these were selected as important predictors of disease-free survival in their models [44]. In one study, Vujasinovic et al. showed that a model incorporating GLCM in H&E and pankeratin-stained biopsies along with other clinicopathological variables could predict distant metastasis with up to 90% accuracy [45]. Interestingly, it appears that there are some similarities but important differences in the features identified as meaningful predictors of response to NAC compared to those useful for long-term prognosis in the non-neoadjuvant setting. Despite one of these studies showing utility of GLCM, the other studies did not find nuclear intensity or similar texture-based features to be significant predictors. Similarly, in the previously mentioned study by Veta et al., they showed nuclear area as a prognosticator of patient survival, whereas we did not find nuclear area a significant predictor of NAC response [40]. These differences can likely be attributed to the different biological phenomenon being studied, i.e., acute response to chemotherapy versus long-term tumor behavior. Another possibility is that different biological types of breast cancer may each have unique predictive features. In our study, we evaluated only HER2-positive and triple-negative tumors whereas many of these other studies looked at a significant number of ER-positive tumors.

The current study has important limitations. The study was retrospective in nature and limited by a relatively small sample size. Increased sample size may have allowed for detection of additional important predictors that did not reach statistical significance in this study. In particular, we did not see significant differences due to receptor status as there were few patients in each group when broken down into receptor subtypes, and we did not include ER-positive/HER2-negative tumors since none of these cases in our cohort had achieved a pCR. It would be informative to know how the predictive features behaved by receptor subtype if there was sufficient sample size to allow for this type of analysis. In this study, we evaluated only a small number of pre-selected nuclear features, which means important predictive features may yet to be uncovered. This does, however, have the advantage of decreasing risk of false-positive findings and over fitting the data, thus, improving the likelihood of being able to reproduce our findings. We also only evaluated tumor nuclear features in this study. Given the previous research that has identified significant predictors of response to NAC, our models could likely be improved by including evaluation of TILs, mitotic activity, cytoplasmic features, necrosis, and stromal features. At last, our study did not include an external validation set to test the reproducibility of nuclear intensity and GLCM-COR to predict the response to NAC, and thus, the findings will need to be verified in an independent cohort.

In conclusion, our study describes using an AI platform on breast core needle biopsies to predict the response of high-risk breast cancers to NAC. This provides the basis for the further development of prediction tools that are both rapid and reproducible, which could be integrated into the pathologist's workflow. Overall, these novel findings have the potential to improve clinical care and outcomes for patients with high-risk and advanced breast cancer.

Acknowledgments The authors wish to thank the Breast Site Group at Sunnybrook Health Sciences Centre for their continued support and intellectual discussions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by DWD, AL, and ST. The first draft of the manuscript was written by DWD and AL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Dr. Lu, Dr. Sadeghi-Naini, and Dr. Tran received funding from the tri-council (Government of Canada) New Frontiers in Research Fund. Dr. Tran lab is funded in part, by the Terry Fox Research Institute and the Women's Health Golf Classic Foundation Fund. Dr. Sadeghi-Naini holds a York Research Chair in Quantitative Imaging and Smart Biomarkers, and received funding from the Natural

Sciences and Engineering Research Council of Canada and the Terry Fox Research Institute.

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

Code availability Code generated during the current study is available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

Ethical approval This study was approved by the institutional research ethics board.

Informed consent Waived as per the institutional research ethics board.

References

1. Untch M, Konecny GE, Paepke S et al (2014) Current and future role of neoadjuvant therapy for breast cancer. Breast 23:526–537
2. Murphy BL, Day CN, Hoskin TL et al (2018) Neoadjuvant chemotherapy use in breast cancer is greatest in excellent responders: triple-negative and HER2+ subtypes. Ann Surg Oncol 25:2241–2248
3. von Minckwitz G, Untch M, Blohmer JU et al (2012) Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 30:1796–1804
4. Cortazar P, Zhang L, Untch M et al (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 384:164–172
5. Spring LM, Fell G, Arfe A et al (2020) Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis. Clin Cancer Res 26:2838–2848
6. Kim KI, Lee KH, Kim TR et al (2014) Ki-67 as a predictor of response to neoadjuvant chemotherapy in breast cancer patients. J Breast Cancer 17:40–46
7. Tan QX, Qin QH, Yang WP et al (2014) Prognostic value of Ki67 expression in HR-negative breast cancer before and after neoadjuvant chemotherapy. Int J Clin Exp Pathol 7:6862–6870
8. Erbes T, Stickeler E, Rücker G et al (2016) BMI and pathologic complete response to neoadjuvant chemotherapy in breast cancer: a study and meta-analysis. Clin Breast Cancer 16:e119–e132
9. Matsubara N, Mukai H, Fujii S et al (2013) Different prognostic significance of Ki-67 change between pre- and post-neoadjuvant chemotherapy in various subtypes of breast cancer. Breast Cancer Res Treat 137:203–212
10. Jung YY, Hyun CL, Jin MS et al (2016) Histomorphological factors predicting the response to neoadjuvant chemotherapy in triple-negative breast cancer. J Breast Cancer 19:261–267
11. Ono M, Tsuda H, Shimizu C et al (2012) Tumor-infiltrating lymphocytes are correlated with response to neoadjuvant chemotherapy in triple-negative breast cancer. Breast Cancer Res Treat 132:793–805
12. Asano Y, Kashiwagi S, Goto W et al (2018) Prediction of treatment response to neoadjuvant chemotherapy in breast cancer by subtype using tumor-infiltrating lymphocytes. Anticancer Res 38:2311–2321
13. Kraus JA, Beriwal S, Dabbs DJ et al (2012) Predictors of pathologic complete response after standard neoadjuvant chemotherapy in triple-negative breast carcinoma. Appl Immunohistochem Mol Morphol 20:334–339
14. Li X, Kanbour-Shakir A, Dabbs DJ et al (2013) Morphologic features do not influence response to trastuzumab-containing neoadjuvant chemotherapy in HER2-positive breast cancer. Appl Immunohistochem Mol Morphol 21:420–425
15. Tran WT, Jerzak K, Lu FI et al (2019) Personalized breast cancer treatments using artificial intelligence in radiomics and pathomics. J Med Imaging Radiat Sci 50:S32–S41
16. Ibrahim A, Gamble P, Jaroensri R et al (2020) Artificial intelligence in digital breast pathology: techniques and applications. Breast 49:267–273
17. Bera K, Schalper KA, Rimm DL et al (2019) Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 16:703–715
18. Landmann A, Farrugia DJ, Zhu L et al (2018) Low estrogen receptor (ER)-positive breast cancer and neoadjuvant systemic chemotherapy: is response similar to typical ER-positive or ER-negative disease? Am J Clin Pathol 150:34–42
19. Prabhu JS, Korlimarla A, Desai K et al (2014) A majority of low (1-10%) ER positive breast cancers behave like hormone receptor negative tumors. J Cancer 5:156–165
20. Symmans WF, Peintinger F, Hatzis C et al (2007) Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 25:4414–4422
21. Vahadane A, Peng T, Sethi A et al (2016) Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans Med Imaging 35:1962–1971
22. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Third international conference on learning representations, ICLR 2015—conference track proceedings, pp 1–14
23. Deng J, Dong W, Socher R et al (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Miami, FL, pp 248–255
24. Nair V, Hinton EG (2010) Rectified linear units improve restricted Boltzmann machines. In: ICML'10: Proceedings of the 27th international conference on international conference on machine learning. https://dl.acm.org/doi/10.5555/3104322.3104425. Accessed 12 Oct 2020
25. Lin M, Chen Q, Yan S (2014) Network in network. In: Second international conference on learning representations, ICLR 2014—conference track proceedings, pp 1–10
26. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge
27. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. Comput Vis Pattern Recogn 9351:234–241
28. Dong H, Yang G, Liu F et al (2017) Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. Commun Comput Inf Sci 723:506–517
29. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition, pp 770–778
30. Kumar N, Verma R, Anand D et al (2020) A multi-organ nucleus segmentation challenge. IEEE Trans Med Imaging 39:1380–1391
31. Kumar N, Verma R, Sharma S et al (2017) A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans Med Imaging 36:1550–1560
32. Kumar N, Verma R, Sharma S et al (2018) Multi-organ nucleus segmentation challenge. In: International conference on medical

image computing and computer-assisted intervention. https://monuseg.grand-challenge.org/Data. Accessed 12 Oct 2020
33. Bloice MD, Roth PM, Holzinger A (2019) Biomedical image augmentation using augmentor. Bioinformatics 35:4522–4524
34. Gutman DA, Khalilia M, Lee S et al (2017) The digital slide archive: a software platform for management, integration, and analysis of histology for cancer research. Cancer Res 77:e75–e78
35. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3:610–621
36. Ataseven B, Lederer B, Blohmer JU et al (2015) Impact of multifocal or multicentric disease on surgery and locoregional, distant and overall survival of 6,134 breast cancer patients treated with neoadjuvant chemotherapy. Ann Surg Oncol 22:1118–1127
37. Romo-Bucheli D, Janowczyk A, Gilmore H et al (2016) Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images. Sci Rep 6:32706
38. Cireşan DC, Giusti A, Gambardella LM et al (2013) Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv 16:411–418
39. Wang H, Cruz-Roa A, Basavanhally A et al (2014) Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J Med Imaging (Bellingham) 1:034003
40. Veta M, Kornegoor R, Huisman A et al (2012) Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer. Mod Pathol 25:1559–1565
41. Whitney J, Corredor G, Janowczyk A et al (2018) Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. BMC Cancer 18:610
42. Adur J, Carvalho HF, Cesar CL et al (2014) Nonlinear optical microscopy signal processing strategies in cancer. Cancer Inform 13:67–76
43. Lu C, Romo-Bucheli D, Wang X et al (2018) Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. Lab Investig 98:1438–1448
44. Chen JM, Qu AP, Wang LW et al (2015) New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images. Sci Rep 5:10690
45. Vujasinovic T, Pribic J, Kanjer K et al (2015) Gray-level co-occurrence matrix texture analysis of breast tumor images in prognosis of distant metastatis risk. Microsc Microanal 21:646–654

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