The Breast 66 (2022) 183–190

Contents lists available at ScienceDirect

The Breast

journal homepage: www.journals.elsevier.com/the-breast

Deep learning with biopsy whole slide images for pretreatment prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer : A multicenter study

Bao Li [a,b,1], Fengling Li [c,d,1], Zhenyu Liu [b,h,1], FangPing Xu [e,1], Guolin Ye [f], Wei Li [f], Yimin Zhang [g], Teng Zhu [i], Lizhi Shao [b], Chi Chen [b,j], Caixia Sun [b,j], Bensheng Qiu [a], Hong Bu [c,d,***], Kun Wang [i,**], Jie Tian [a,b,j,k,*]

[a] Center for Biomedical Imaging, University of Science and Technology of China, Hefei, 230026, China
[b] CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[c] Department of Pathology, West China Hospital, Sichuan University, Chengdu, 610041, China
[d] Institute of Clinical Pathology, West China Hospital, Sichuan University, Chengdu, 610041, China
[e] Department of Pathology, Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, Guangzhou, 510080, China
[f] The First People's Hospital of Foshan, Foshan, 528000, China
[g] Diagnosis & Treatment Center of Breast Diseases, Clinical Research Center, Shantou Central Hospital, Shantou, 515000, China
[h] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100080, China
[i] Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, Guangzhou, 510080, China
[j] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, 100191, China
[k] Key Laboratory of Big Data-Based Precision Medicine, Beihang University, Ministry of Industry and Information Technology, People's Republic of China, Beijing, 100191, China

ARTICLE INFO

Keywords:
Breast cancer
Neoadjuvant chemotherapy
Pathological complete response
Whole-slide image
Deep learning

ABSTRACT

Introduction: Predicting pathological complete response (pCR) for patients receiving neoadjuvant chemotherapy (NAC) is crucial in establishing individualized treatment. Whole-slide images (WSIs) of tumor tissues reflect the histopathologic information of the tumor, which is important for therapeutic response effectiveness. In this study, we aimed to investigate whether predictive information for pCR could be detected from WSIs.

Materials and methods: We retrospectively collected data from four cohorts of 874 patients diagnosed with biopsy-proven breast cancer. A deep learning pathological model (DLPM) was constructed to predict pCR using biopsy WSIs in the primary cohort, and it was then validated in three external cohorts. The DLPM could generate a deep learning pathological score (DLPs) for each patient; stromal tumor-infiltrating lymphocytes (TILs) were selected for comparison with DLPs.

Results: The WSI feature-based DLPM showed good predictive performance with the highest area under the curve (AUC) of 0.72 among the cohorts. Alternatively, the combination of the DLPM and clinical characteristics offered a better prediction performance (AUC >0.70) in all cohorts. We also evaluated the performance of DLPM in three different breast subtypes with the best prediction for the triple-negative breast cancer (TNBC) subtype (AUC: 0.73). Moreover, DLPM combined with clinical characteristics and stromal TILs achieved the highest AUC in the primary cohort (AUC: 0.82) and validation cohort 1 (AUC: 0.80).

Conclusion: Our study suggested that WSIs integrated with deep learning could potentially predict pCR to NAC in breast cancer. The predictive performance will be improved by combining clinical characteristics. DLPs from DLPM can provide more information compared to stromal TILs for pCR prediction.

* Corresponding author. Center for Biomedical Imaging, University of Science and Technology of China, Hefei, 230026, China.
** Corresponding author. Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, Guangzhou 510080, China
*** Corresponding author. Department of Pathology, West China Hospital, Sichuan University, Chengdu, 610041, China.
E-mail addresses: hongbu@scu.edu.cn (H. Bu), wangkun@gdph.org.cn (K. Wang), jie.tian@ia.ac.cn (J. Tian).
[1] First authors contributed equally.

https://doi.org/10.1016/j.breast.2022.10.004
Received 9 November 2021; Received in revised form 18 September 2022; Accepted 11 October 2022
Available online 19 October 2022

1. Introduction

Breast cancer is a malignant tumor affecting females with the highest incidence worldwide [1]. Neoadjuvant chemotherapy (NAC) has been guided as the standard care for patients with breast cancer [2]. Ideally, the response to NAC is described by the pathological complete response (pCR), which is associated with favorable outcomes [3]. Thus, the prediction of pCR before the treatment is crucial to stratify patients early and guide the subsequent treatment regimen.

Molecular subtype and clinical staging are important indicators for the implementation of NAC [3]. These factors, such as the tumor size [4] and immunohistochemistry markers of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor 2 (HER2) and Ki67 index, were associated with pCR [5]. In addition, some models have been proposed for predicting the pCR to NAC from different prospects, including clinic-pathological variables [6,7], genomic signatures [8], and even radiomic features from pretreatment [9] and mid-treatment [10] magnetic resonance imaging (MRI). Although these predictive factors showed potential in predicting response to NAC, developing novel predictors was still essential for clinical management. Microscopic examination of tissue stained with hematoxylin and eosin (H&E) is one of the standard diagnostic procedures for breast cancer. These stained tissues can be digitally scanned as whole-slide images (WSIs). However, few studies have attempted to use WSIs to predict therapeutic response in breast cancer using state-of-the-art techniques.

WSIs for tumor tissues can disclose microscopic histopathologic information of the tumor that can be highly associated with patients' prognosis and therapeutic response [5,11]. Several studies have demonstrated that the extracted quantitative features from WSIs are associated with the therapy response. The WSI's spatial information, such as stromal tumor-infiltrating lymphocytes (TILs) density, has been reported to be associated with the therapeutic response and prognosis [12,13]. Some studies have also shown that the morphological features extracted from WSI could predict therapeutic responses in rectal cancer [14,15]. On the other hand, the distribution statistics of WSI patch-level features could be aggregated to patient-level features to predict clear cell renal cell carcinoma prognosis [16]. These earlier studies guided us to investigate pretreatment pCR prediction based on WSIs in breast cancer.

In recent years, deep neural network (DNN) has shown superior performance in learning features from medical images. Previously, the procedures for morphological feature extraction for whole-slide analysis were time-consuming, and the annotation of the region of interest (ROI) for WSIs required manual efforts. However, with DNN, high dimension spatial information of WSI could be automatically captured and represented. Earlier studies of WSI that focused on quantitative analysis using DNN have shown promising results in survival prediction in hepatocellular carcinoma resection [17], outcome prediction of colorectal cancer [18], and pathomic fusion for cancer diagnosis and prognosis [19]. These achievements in deep learning have motivated us in developing a DNN-based model for high-level image feature representation without involving pathologists' annotation.

In this multicenter retrospective study, we aimed to develop a deep learning pathological model (DLPM) using WSIs in breast cancer and validate whether it could predict pCR to NAC. We also investigate whether the combination of DLPM with clinical information would improve prediction performance. In addition, a comparison was performed between the stromal TILs and deep learning pathological scores (DLPs) generated by DLPM.

2. Materials and methods

2.1. Patients

Our study was approved by the West China Hospital of Sichuan University Ethics Committee (2021–764) and Guangdong Provincial People's Hospital Ethics Committee (GDREC2019800H). The written informed consent of patients was waived as this was a retrospective study. In total, 874 patients diagnosed with biopsy-proven breast cancer were collected between January 2015 and December 2019 from four cohorts. The cohort with the largest number of patients was assigned as the primary cohort (PC), and the other three hospitals were used as external validation cohorts (VC1-VC3) (Supplementary Section 1 and 2).

2.2. The acquisition and preprocessing of WSIs

H&E-stained slides from formalin-fixed paraffin-embedded (FFPE) of pretreatment needle biopsies, were requested within 1–4 weeks before NAC. All slides were digitized for further analysis. WSIs, each having approximately 10 gigapixels, were preprocessed before further analysis. The preprocessing procedure involved three main steps: foreground mask extraction, image tiling, and staining normalization (Supplementary Section 3 and 4).

2.3. DLPM construction based on ResNet18 using WSIs

In this study, we hypothesized that only cell-predominant patches within the WSI contained predictive information of treatment response to NAC. Herein, we first built a patch-level classifier based on pretrained ResNet18 [20] to determine each patch's category. For the patches classified as cancer cell predominant patches or stromal cell predominant patches, 512-dimension output was extracted from the last convolutional layer of ResNet18 as the patch-level feature. Then, the patch-level features of each patient were combined to generate a patient-level feature by computing the 6 statistics (mean, standard variation, max, median, kurtosis, skew) along the patch axis [16].

After the feature aggregation and selection, 11 features remained for the primary cohort. Multivariate logistic regression with 10-fold cross-validation was performed on the primary cohort to build DLPM for pCR prediction. The output of DLPM was then used to generate DLPs for each patient. To gather information from clinical characteristics, age, tumor stage (cT), lymph node stage (cN), ER/PR/HER2 status, and Ki67 status were selected to build the solely clinical model, named multivariate logistic regression clinical model (MLCM). MLCM was also trained in the PC with 10-fold cross-validation. We then combined the DLPM and MLCM to build up a deep learning pathological clinical model (DLPCM) (Supplementary Section 5). The output of DLPCM was referred to deep learning pathological clinical score (DLPCs).

To assess the quantitative performance of DLPM, MLCM, and DLPCM, the receiver operating characteristic (ROC) curves and the AUC was computed.

2.4. DLPM performance in different breast cancer subtypes

According to ASCO/CAP guidelines [21], nuclear staining of ER/PR by immunohistochemistry (IHC) with $<1\%$ positive tumor cells was defined as ER/PR negative, while that with $\geq 1\%$ positive tumor cells was defined as ER/PR positive. ER and/or PR was also referred as Hormone receptors (HR). The cutoffs of the Ki67 index were $<20\%$ for low expression and $\geq 20\%$ for high expression [22]. HER2 positive was determined with IHC 3+ or IHC 2+ with amplification by *in situ* hybridization (ISH), while HER2 negative was defined as IHC 0/1+ or IHC 2+ without amplification by ISH [23]. Subtypes of breast cancer were significantly variant in the standard treatment and prognosis [24]. The patients from our four cohorts were divided into three molecular subtypes based on the receptor status of primary tumors: HR positive and HER2 negative (HR+ and HER2-), HER2 positive (HER2+), and ER, PR, HER2 negative (TNBC). The AUC of ROC was chosen for the DLPM model quantitative performance evaluation in three subgroups.

2.5. Comparison between DLPs and TILs

We selected stromal TILs for comparison with DLPs. The stromal TILs

were a predictor of the response to NAC therapy in breast cancer patients [25]. The assessment of TILs scores was conducted in the PC and VC1, using the international TILs working group guidelines [26]. The comparison between DLPs and stromal TILs was performed in combination with the clinical score. Then the two new scores were combined to evaluate the overall performance (Supplementary Section 6). We also categorized TILs into three predetermined groups following Denkert et al. [12]: low subgroups (≤10% TILs), intermediate subgroups (11–59%), and high subgroups (≥60% TILs).

2.6. Statistics

The age was treated as a continuous variable and summarized as mean ± standard deviation. It was compared between pCR and non-pCR patients by the Mann-Whitney *U* test. Differences in ER, PR, HER2, Ki67, cT, cN and histological grade between patients with and without pCR were tested using the Chi-square test or Fisher exact test among all cohorts. All statistical tests were two-sided and significant at $P < 0.05$. Statistical analyses were performed in python (version 3.6.10). The 95% confidence interval (CI) was used for AUC. This analysis was performed on R software (version March 1, 1093).

3. Results

3.1. Patient characteristics

The clinical characteristics of the patients were summarized in Table 1. The results showed that the pCR rate across the four cohorts was variant, with 17.2%, 41.4%, 29.6%, and 27.3% in PC, VC1, VC2, and VC3, respectively. Regarding age, we did not observe a significant difference ($P > 0.05$) between patients with and without pCR for all cohorts. Meanwhile, the cT characteristic in VC1 differed between pCR and non-pCR patients, whereas such a difference in cN characteristic was only observed in VC3. A significant difference was detected in ER, PR, HER2 status and stromal TILs concentration in both PC and VC1. As for VC2, only ER and HER2 status were associated with pCR. However, the cohort with limited patients like VC3 showed no difference for the three clinical molecular characteristics. Furthermore, the Ki67 index was associated with pCR in the PC only. Meanwhile, we observed that histological grade in PC and VC2 differed between pCR and non-pCR patients but not in VC1 and VC3. These results suggested that the status of ER, PR, HER2, and stromal TILs were correlated with pCR, especially in large cohorts like PC and VC1.

3.2. Prediction performance of models for pCR

The ResNet18-patch classifier demonstrated an efficient automatic detection of the cell predominant patches labeled red or light red (Fig. 1). After feature extraction, aggregation, and selection, 11 deep learning pathological features were used for the DLPM construction. The DLPM showed reliable performance on pCR prediction with PC and external VCs (Fig. 2, Table 2). Specifically, the AUC of DLPM was 0.72 (95% CI 0.66–0.78) in the PC, while the best AUC value among the external VCs reached 0.71 (95% CI 0.57–0.85). On the other hand, the clinical model also achieved good performance in PC and VC1, with AUC values of 0.79 (95% CI 0.74–0.84) and 0.76 (95% CI 0.70–0.82) respectively. The MLCM achieved poor performance in VC2 and VC3 properly due to the lack of enrolled patients.

Table 1
Clinical characteristics of patients in the primary and validation cohorts.

Characteristic	Primary cohort		Validation cohort 1		Validation cohort 2		Validation cohort 3	
	(N = 482)		(N = 222)		(N = 115)		(N = 55)	
	N (pCR rate%)	*P*-value	N (pCR rate%)	*P*-value	N (pCR rate%)	*P*-value	N (pCR rate%)	*P*-value
Age (years) mean ± SD	48.4 ± 10.1	0.212	48.7 ± 9.3	0.160	49.9 ± 9.3	0.480	49.6 ± 11.2	0.200
ER		<0.001*		<0.001*		0.046*		0.847
Positive	308 (9.7%)		143 (30.1%)		78 (23.1%)		25 (28.0%)	
Negative	174 (30.5%)		79 (62.0%)		37 (43.2%)		30 (26.7%)	
PR		<0.001*		<0.001*		0.495		0.978
Positive	286 (9.4%)		128 (31.3%)		78 (26.9%)		20 (30.0%)	
Negative	196 (28.6%)		94 (55.3%)		37 (35.1%)		35 (25.7%)	
HER2		<0.001*		<0.001*		<0.001*		0.172
Positive	137 (38.0%)		87 (63.2%)		40 (50.0%)		23 (39.1%)	
Negative	345 (9.0%)		135 (27.4%)		75 (18.7%)		32 (18.8%)	
Ki67		0.013*		0.488		0.879		0.886
Positive	357 (19.9%)		165 (43.0%)		92 (30.4%)		50 (28.0%)	
Negative	125 (9.6%)		57 (36.8%)		23 (26.1%)		5 (20.0%)	
cT stage		0.253		<0.001*		0.661		0.284
cT1	18 (27.8%)		15 (46.7%)		9 (33.3%)		0 (0.0%)	
cT2	193 (16.6%)		170 (47.6%)		85 (31.8%)		14 (42.9%)	
cT3	90 (22.2%)		30 (13.3%)		8 (25.0%)		30 (20.0%)	
cT4	181 (14.4%)		7 (0.0%)		13 (15.4%)		11 (27.3%)	
cN stage		0.285		0.426		0.815		0.046*
cN0	37 (18.9%)		87 (48.3%)		15 (20.0%)		7 (28.6%)	
cN1	207 (20.8%)		117 (36.8%)		79 (30.4%)		19 (10.5%)	
cN2	101 (13.9%)		13 (38.5%)		14 (35.7%)		21 (47.6%)	
cN3	137 (13.9%)		5 (40.0%)		7 (28.6%)		8 (12.5%)	
Histological grade		<0.001*		0.054		0.023*		0.446
Grade 1, 2	380 (13.9%)		110 (34.5%)		78 (24.4%)		24 (4.1%)	
Grade 3	102 (29.4%)		112 (48.2%)		28 (50%)		6 (0%)	
Unknown	0 (0%)		0 (0%)		9 (11.1%)		25 (56%)	
Stromal TILs		<0.001*		0.007*		–	–	–
Low (0–10%)	372 (12.6%)		115 (32.2%)		–		–	
Intermediate (11–59%)	103 (31.1%)		90 (48.9%)		–		–	
High (≥60%)	7 (57.1%)		17 (64.7%)		–		–	

Note: Data are mean (SD) or N (pCR rate%). Chi-square ($\chi 2$) tests were used to test whether the variable composition varied significantly between pCR and non-pCR patients. *P*-value<0.05 indicates that the variable distribution varied significantly between pCR and non-pCR patients. pCR = pathological complete response. non-pCR = non-pathological complete response.

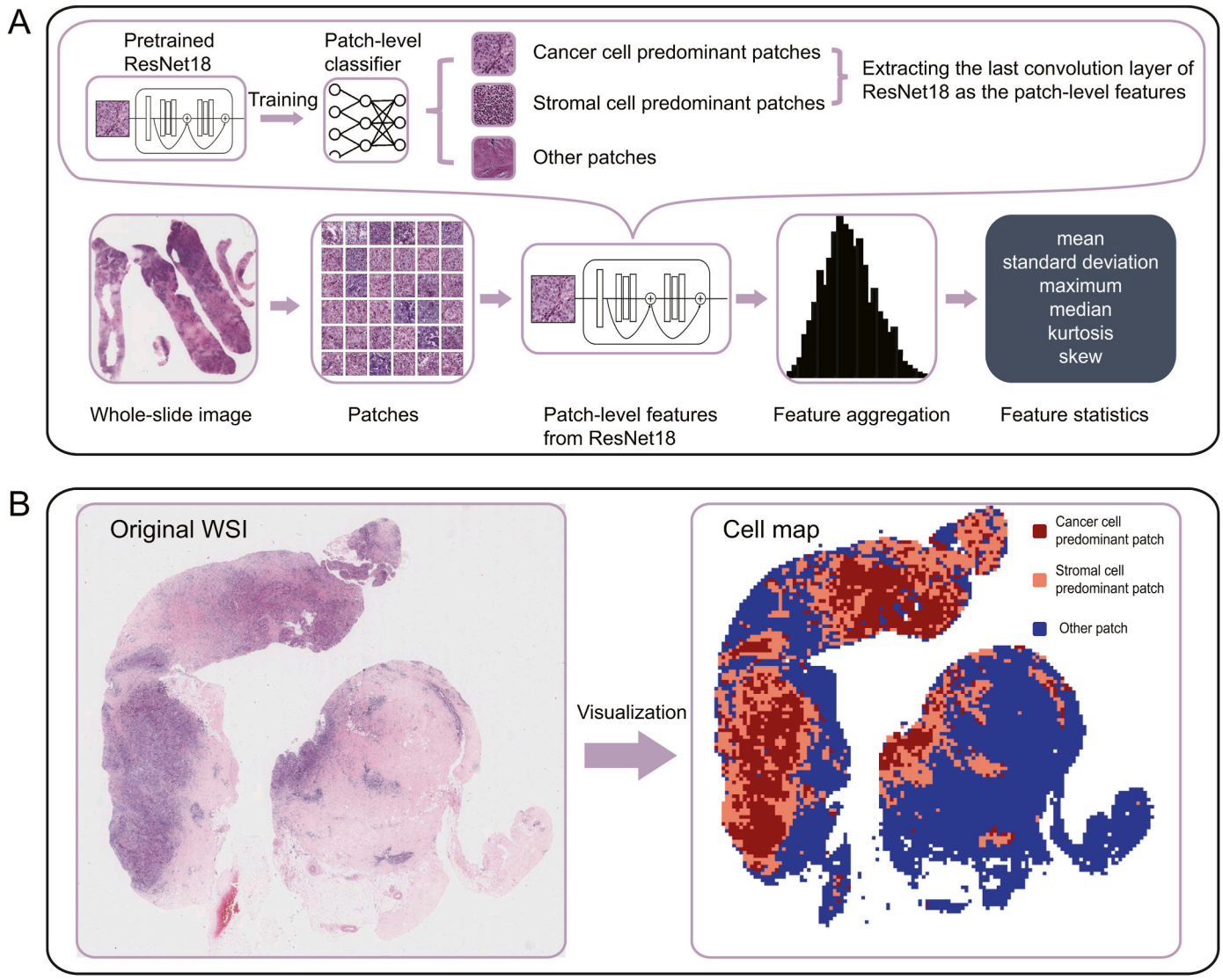

Fig. 1. Schematic of Deep Learning Pathological Feature Extraction. (A) The automatic deep learning pathological feature extraction procedure. Patches were tiled with a size of 512 × 512. And a patch-level classifier was trained using transfer learning with ResNet18. For the patches classified as the cancer cell or stromal cell predominant patches, the patch-level feature was extracted from the last convolutional layer output of the patch classifier. The patient's patch-level features from all WSIs were then aggregated to create patient-level features. (B) Left: The patient's original WSI. Right: visualization of labeled patches. Cancer cell predominant patches are labeled in red. Stromal cell predominant patches are labeled in light red. Other patches are labeled in blue. Background patches are labeled in white. WSI = whole-slide image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The combination of MLCM with DLPM offered a better prediction performance in all cohorts (Fig. 2). Notably, the AUC of the DLPCM was close to 0.8 in both the PC and VC1. Possibly due to a limited number of patients enrolled, the prediction performance in VC2 and VC3 was not as good as that in PC and VC1. The case was more unpleasant in VC3, in which the sample size (55 patients) was the lowest. Even with this limited sample size, the VC2 and VC3 data yield AUC >0.70 in DLPCs.

We also divided the enrolled patients into three subgroups based on receptor status: HR+ and HER2- (453 patients), HER2+ (287 patients), and TNBC (134 patients). Of these subgroups, TNBC had the highest performance of DLPM with an AUC of 0.73 (95% CI, 0.63–0.83). Fig. 3 shows the ROC curves and corresponding AUCs of each subgroup.

3.3. Comparison between DLPs and TILs

To further determine the predictive value of DLPs, the stromal TILs score was used for the comparison. TILs, as a continuous variable, showed its predictive ability for pCR in PC (AUC: 0.66; Fig. S4A), VC1 (AUC: 0.63; Fig. S4B) and among all subtypes (AUC; HR+ and HER2-: 0.69, HER2+: 0.63, TNBC: 0.57; Fig S5). When the TILs were categorized into three predefined groups, TILs were significantly associated with pCR in all breast subtypes (Table S3) and higher TILs were related to a higher pCR rate among all subtypes (Fig. S6). With the MLCM as the baseline model, we evaluated the incremental value with the inclusion of DLPs or TILs. We observed only a slight improvement in VC1 (AUC, 0.77 vs 0.76; not significant) when predicting pCR using MLCM with TILs score. Nevertheless, the MLCM with DLPs achieved an AUC of 0.81 in PC and 0.79 in VC1. These results (AUC of MLCM with DLPs vs MLCM with TILs; 0.81 vs 0.79, in PC; 0.79 vs 0.77, in VC1) suggested that the WSIs' DLPs were different from TILs and that DLPs could provide more supplementary information for clinical signatures about the pCR prediction than the TILs score. Under the combination of all these scores (MLCM with TILs and DLPs), the highest performance score was achieved with an AUC of 0.82 in PC and 0.80 in VC1, suggesting a lower contribution of TILs than DLPs in predicting pCR (Fig. 4). In the univariate and multivariate analyses, TILs showed significance in both PC

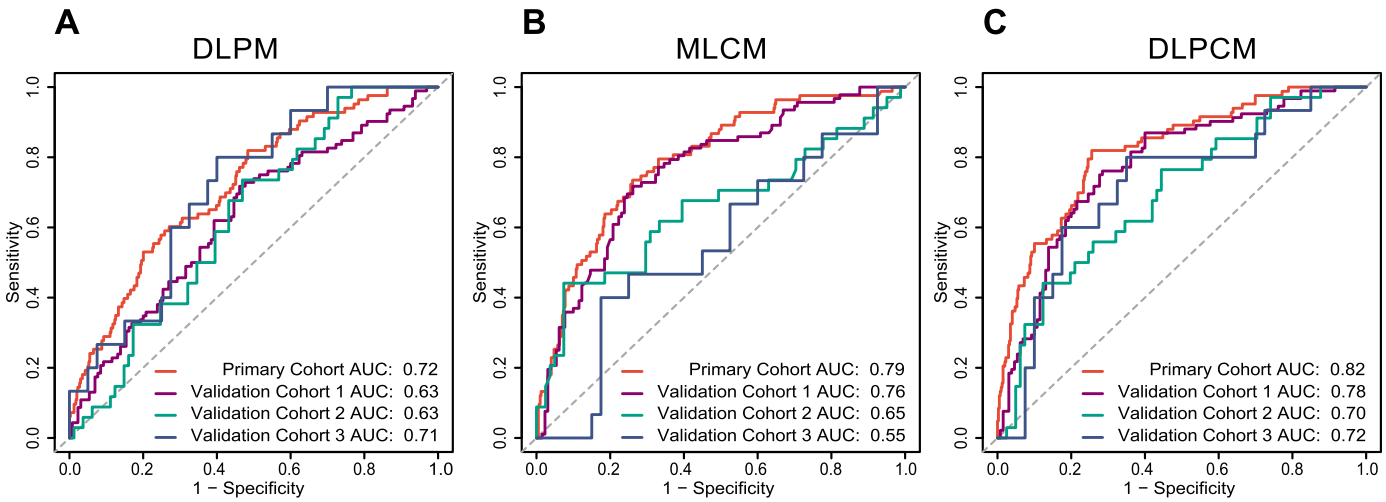

Fig. 2. The ROC curve for prediction performance in the (A) DLPM, (B) MLCM, and combined model (C) DLPCM among all cohorts. ROC = receiver operating characteristic. AUC = area under the receiver operating characteristic curve. DLPM = deep learning pathological model. MLCM = multivariate logistic regression clinical model. DLPCM = deep learning pathological clinical model.

Table 2
The predictive performance of models for pCR across cohorts.

Cohort	DLPM AUC (95% CI)	MLCM AUC (95% CI)	DLPCM AUC (95% CI)
Primary Cohort	0.72 (0.66–0.78)	0.79 (0.74–0.84)	0.82 (0.77–0.87)
Validation Cohort 1	0.63 (0.56–0.70)	0.76 (0.70–0.82)	0.78 (0.72–0.84)
Validation Cohort 2	0.63 (0.53–0.73)	0.65 (0.53–0.77)	0.70 (0.60–0.80)
Validation Cohort 3	0.71 (0.57–0.85)	0.55 (0.38–0.72)	0.72 (0.56–0.88)

Note: pCR = pathological complete response. DLPM = deep learning pathological model. MLCM = multivariate logistic regression clinical model. DLPCM = deep learning pathological clinical model. AUC = area under the curve. CI = confidence interval.

(Table S4) and VC1 (Table S5). When DLPs were included in the multivariate analysis, TILs were no longer significant in the PC. However, DLPs remained statistically significant in both PC and VC1, indicating that DLPs contained predictive information from TILs and even outperformed it.

3.4. Visualization of DLPM

We used the Full-Grad [27] visualization approach to better interpret the findings obtained by the WSI feature-based DLPM (Supplementary Section 7). We randomly selected all the cell-predominant patches of three patients with pCR and three patients with non-pCR. The Full-Grad then generated an overlaid attention map for each input tile. For the visualization results, we frequently observed lymphocyte infiltration and aggregation (Fig. S7A), vacuoles around the tumor cells (Fig. S7B), the presence of mitosis (Fig. S7C), and pleomorphic nuclei with multiple atypical nucleoli (Fig. S7D), which contributed to the DLPM prediction.

4. Discussion

In this study, we have shown that DLPM constructed from WSIs could potentially predict response in patients with breast cancer. In addition,

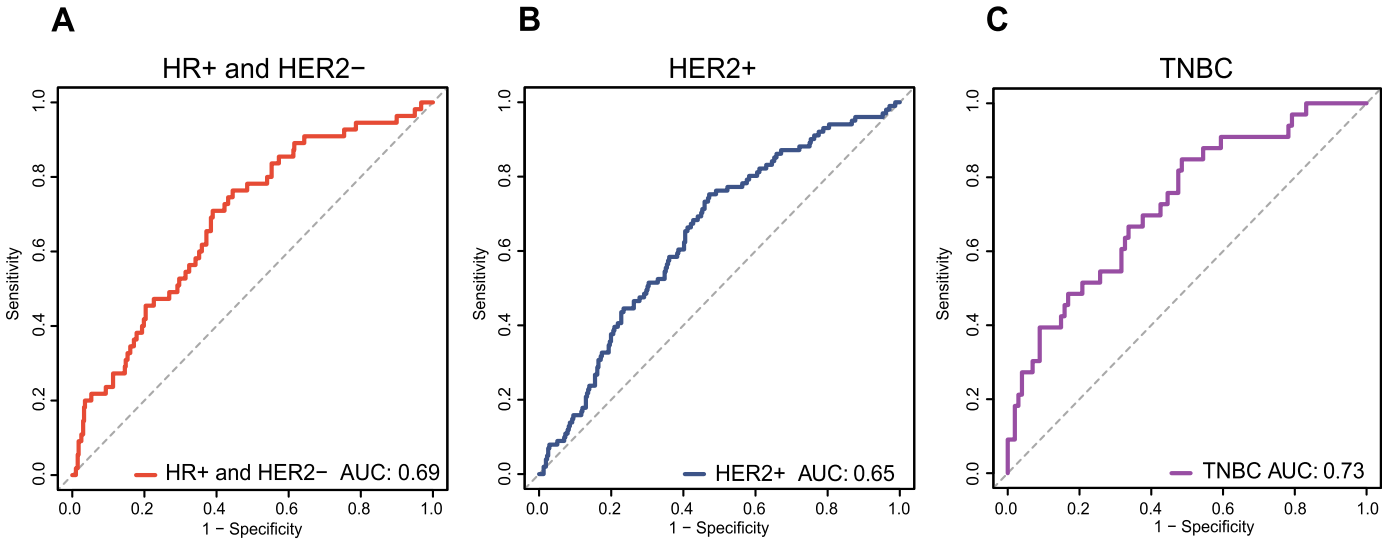

Fig. 3. The ROC curve for prediction performance using DLPM in the (A) HR+ and HER2-, (B) HER2+, and (C) TNBC subgroup. ROC = receiver operating characteristic. AUC = area under the receiver operating characteristic curve. DLPM = deep learning pathological model. HR = hormone receptor. HER2 = human epidermal growth factor 2. TNBC = triple-negative breast cancer.

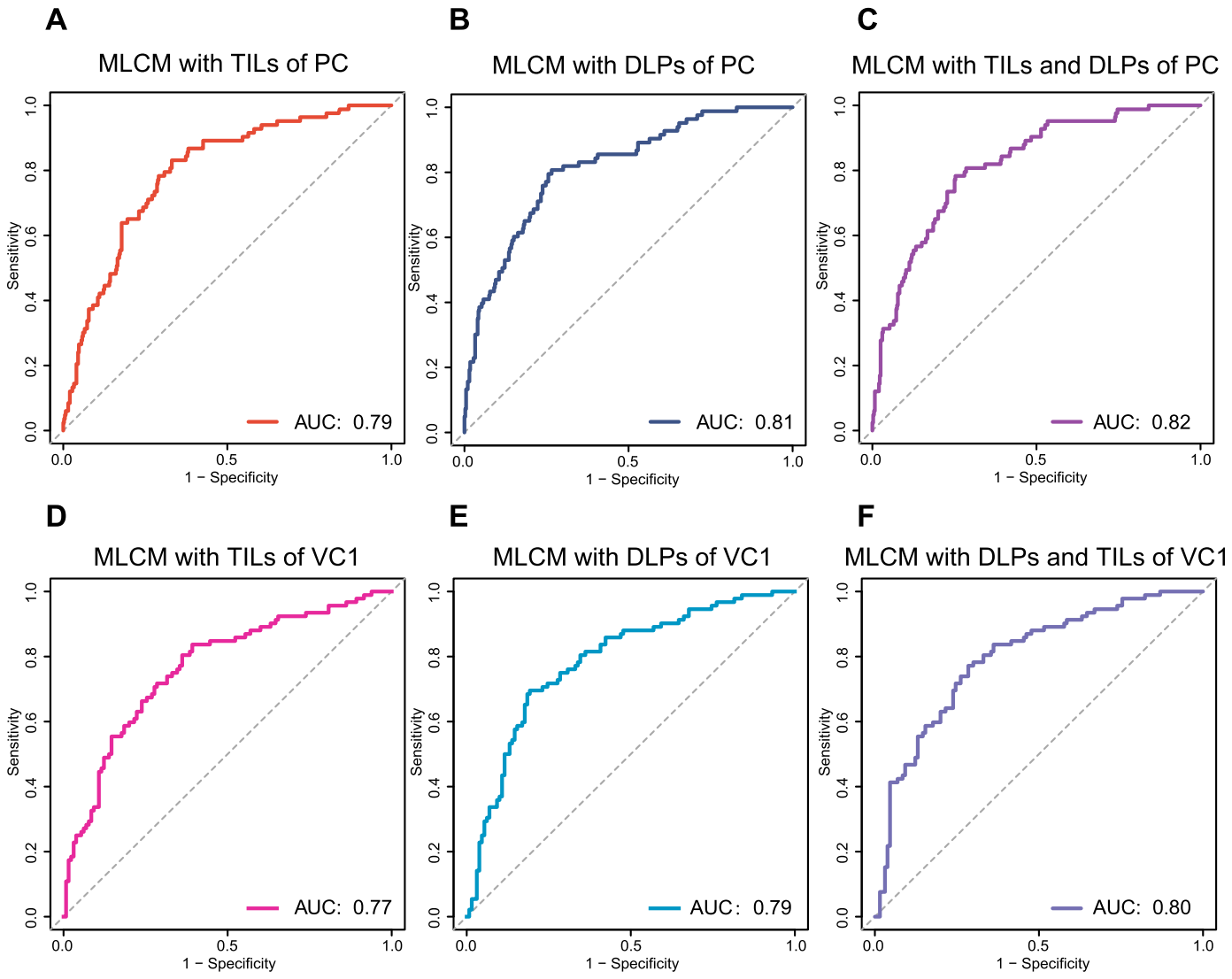

Fig. 4. The ROC curve for prediction performance by combining MLCM with (A) DLPs, (B) TILs, (C) DLPs and TILs in the PC, and (E) DLPs, (F) TILs, (G) DLPs and TILs in the VC1. ROC = receiver operating characteristic. AUC = area under the receiver operating characteristic curve. MLCM = multivariate logistic regression clinical model. DLPs = deep learning pathological score. TILs = tumor-infiltrating lymphocytes score. PC = primary cohort. VC1 = validation cohort 1.

we demonstrated that a combined model, DLPC, which was based on DLPM and MLCM could achieve the best AUC for pCR prediction among all cohorts. We have also seen that patients with TNBC were more likely to have their NAC responses be predicted than the rest subgroups. Moreover, DLPs can provide MLCM with more information regarding pCR prediction than TILs.

The promising performance of DLPM indicated that the deep learning features from WSI were relevant for predicting pCR of NAC in breast cancer, which was also confirmed in external VCs. Also, when DLPM and clinical information were combined, the model, DLPCM, performed best in all cohorts, demonstrating that DLPM might be used as a complement to the model for pretreatment prediction of pCR. The DLPM represented the high-dimension of visually spatial features from WSIs; the stromal TILs score was manually assessed from the stromal region of WSIs, and the ER, PR, HER2, and Ki67 index were molecular features. These three types of information could complement each other in predicting the therapeutic response of cancer treatment. Besides, our findings demonstrated that DLPM had the highest predictive performance for pCR in patients with TNBC, which was consistent with those reported in the previous study [9]. We also explained our model by the attention map generated from Full-Grad [27]. By the attention map, we found that the pCR prediction in our model was associated with histological features of lymphocytes, vacuoles, mitosis, and nuclear atypia. These findings were in accord with the previous results that high histological grade [28] and density of TILs were correlated with pCR [12]. In addition, we found that the presence of vacuoles around the tumor cells may be associated with efficacy, which was not observed before. This finding suggests that AI facilitates the discovery of novel or potential diagnostic features. Therefore, DLPM had the potential to improve pCR prediction and facilitate the better stratification of breast cancer patients in NAC settings, particularly those with TNBC.

Our model selected content area patches and extracted pathological features automatically instead of using manual annotations. This method had several benefits, including reduced feature variances and enhanced feature repeatability. In the preprocessing procedure, patch resizing and staining normalization were performed to reduce the inconsistency from different cohorts or scanners which could improve the feature reproducibility. Furthermore, not only DLPM was proven predictive of pCR in PC but also its results were validated in three external cohorts. Meanwhile, WSIs were available for almost every breast cancer patient. WSI-based quantitative analysis via state-of-the-art methods could facilitate resolving complex clinical matters, with wide prospects for automated cancer analysis systems in the clinic. Although its adaption may seem challenging at present, the potential of

WSIs in pCR prediction is quite promising; and more enthusiastic and creative works need to be devoted to realizing the final breakthrough.

Regarding the comparison of DLPs and TILs, earlier studies have reported that TILs score is associated with the response to NAC in all breast cancer subgroups [12]. Only a few studies report the AUC of single-center pCR prediction using TILs, and these results have not been verified in external cohorts [29,30], with AUCs of 0.645 in the TNBC subgroup with 166 patients and 0.683 in the HER2+ subgroup with 143 patients, respectively. Meanwhile, an AUC of 0.768 was reported in a study of 105 patients with HER2+ breast cancer [31]. These results reveal that the ability of manually assessed TILs to predict pCR varied among cohorts. In addition, according to the ASCO guideline [32], there is no sufficient evidence to support the use of TILs in guiding the decision of neoadjuvant chemotherapy. In most studies, TILs inside the stromal region's area are evaluated since it is a superior and more reproducible parameter [26]. However, TILs mainly contain the information of the stromal lymphocytes and do not involve the tumor cells, while characteristics of the tumor cells such as tumor nuclear features [33] could also be correlated with the response to NAC in breast cancer. Meanwhile, TILs require manual evaluation, resulting in inter/intra-observer variability. With all this in mind, it guides us to explore more high-level information from WSIs associated with therapy efficiency. Alternatively, different from TILs, DLPs is generated automatically from WSI without the need for manual evaluation, implying that DLPs is a good complement to TILs for predicting pCR.

Despite the better performance offered by our study, the limited results of WSIs-based DLPM in VC1 and VC2 may be attributable to patient heterogeneity in breast cancer [34]. Another reason may be the fact that biopsy WSIs are sampled from representative tumor sections but not the whole tumor area. They microscopically reflect the intrinsic characteristics of histopathologic information but lack a macroscopic description of the entire tumor tissue. Medical images, such as MRI [9,35], CT [36], and PET [37], can perform whole tumor area scanning on a macroscopic scale and represent the intra-tumor heterogeneity [35] and the entire tumor environment [38]. Therefore, our model could not achieve the equivalent prediction performance compared to the above modal images and we speculate that the fusion of multiple modal images could lead to a greater breakthrough in pretreatment prediction. Additionally, it is important to note that predicting pCR before NAC is more challenging than after NAC [9], which may also explain our results.

There are some limitations worth addressing in our study. First, although this is a multicenter study, limited numbers of patients were enrolled in our research, especially in VC2 and VC3. More patients who received standard NAC treatment should be recruited in future studies. Second, DLPM in our study was only developed from WSIs reflecting the microscopic scale of histopathologic information. The radiomics signatures can be used to represent the intra-tumor heterogeneity [35] and the entire tumor environment [38]. Hence, MRI, CT, or PET signatures could be combined with DLPM in future investigations. Furthermore, our study was a retrospective study; a well-designed prospective study should be designed to validate the DLPM in the clinical trial to guide NAC treatment.

5. Conclusion

Our study demonstrated that WSI feature-based DLPM could potentially predict pathological complete response to neoadjuvant chemotherapy in breast cancer. When combined with clinical information, the performance of DLPM was improved in all cohorts. Furthermore, our results suggested that the DLPs generated from DLPM can provide more information for pCR prediction than the TILs score.

Data availability statement

The data and clinical information that support the findings of this study could be made available upon reasonable request and with the permission of the corresponding author. The source code for the WSIs tiling, staining normalization, patch classification, feature generation, and model construction has been openly available in the GitHub repository (https://github.com/Boyden/Pretreatment-Prediction-to-NAC-DLPM). A written research agreement and Institutional Review Board's approval should be submitted to define the obligations and manage the risks.

Conflicts of interest

All authors declare no competing interests.

Acknowledgments

This study is supported by grants from the National Key Research and Development Plan of China (No. 2021YFF1201003), the National Natural Science Foundation of China (No. 81922040, 92059103), the Youth Innovation Promotion Association CAS (No. 2019136), Science and Technology Planning Project of Guangzhou City (202002030236), Beijing Medical Award Foundation (YXJL-2020-0941-0758), Science and Technology Special Fund of Guangdong Provincial People's Hospital (No. Y012018218), CSCO-Hengrui Cancer Research Fund (Y-HR2016-067), Guangdong Provincial Department of Education Characteristic Innovation Project (2015KTSCX080), the 1.3.5 Project for Disciplines of Excellence (ZYGD18012), the Technological Innovation Project of Chengdu New Industrial Technology Research Institute (2017-CY02–00026-GX).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2022.10.004.

References

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. Ca - Cancer J Clin 2022;72(1):7–33. https://doi.org/10.3322/caac.21708.
[2] Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast cancer, version 3.2020, NCCN clinical practice guidelines in oncology. J Natl Compr Cancer Netw 2020;18(4):452–78. https://doi.org/10.6004/jnccn.2020.0016.
[3] Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 2014;384(9938):164–72. https://doi.org/10.1016/s0140-6736(13)62422-8.
[4] Goorts B, van Nijnatten TJA, de Munck L, Moossdorff M, Heuts EM, de Boer M, et al. Clinical tumor stage is the most important predictor of pathological complete response rate after neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res Treat 2017;163(1):83–91. https://doi.org/10.1007/s10549-017-4155-2.
[5] Lips EH, Mulder L, de Ronde JJ, Mandjes IA, Koolen BB, Wessels LF, et al. Breast cancer subtyping by immunohistochemistry and histological grade outperforms breast cancer intrinsic subtypes in predicting neoadjuvant chemotherapy response. Breast Cancer Res Treat 2013;140(1):63–71. https://doi.org/10.1007/s10549-013-2620-0.
[6] Pu S, Wang K, Liu Y, Liao X, Chen H, He J, et al. Nomogram-derived prediction of pathologic complete response (pCR) in breast cancer patients treated with neoadjuvant chemotherapy (NCT). BMC Cancer 2020;20(1):1120. https://doi.org/10.1186/s12885-020-07621-7.
[7] Kim SY, Cho N, Choi Y, Lee SH, Ha SM, Kim ES, et al. Factors affecting pathologic complete response following neoadjuvant chemotherapy in breast cancer: development and validation of a predictive nomogram. Radiology 2021;299(2):290–300. https://doi.org/10.1148/radiol.2021203871.
[8] Edlund K, Madjar K, Lebrecht A, Aktas B, Pilch H, Hoffmann G, et al. Gene expression–based prediction of neoadjuvant chemotherapy response in early breast cancer: results of the prospective multicenter EXPRESSION trial. Clin Cancer Res 2021;27(8):2148–58. https://doi.org/10.1158/1078-0432.ccr-20-2662.
[9] Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. Clin Cancer Res 2019;25(12):3538–47. https://doi.org/10.1158/1078-0432.ccr-18-3190.
[10] Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim J-A, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. Radiology 2020;294(1):31–41. https://doi.org/10.1148/radiol.2019182718.

[11] Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 2010;12(4):207. https://doi.org/10.1186/bcr2607.
[12] Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol 2018;19(1):40–50. https://doi.org/10.1016/S1470-2045(17)30904-X.
[13] Sammut SJ, Crispin-Ortuzar M, Chin SF, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature 2022;601(7894):623–9. https://doi.org/10.1038/s41586-021-04278-5.
[14] Shao L, Liu Z, Feng L, Lou X, Li Z, Zhang X-Y, et al. Multiparametric MRI and whole slide image-based pretreatment prediction of pathological response to neoadjuvant chemoradiotherapy in rectal cancer: a multicenter radiopathomic study. Ann Surg Oncol 2020;27(11):4296–306. https://doi.org/10.1245/s10434-020-08659-4.
[15] Zhang F, Yao S, Li Z, Liang C, Zhao K, Huang Y, et al. Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. Clin Transl Med 2020;10(2):e110-e. https://doi.org/10.1002/ctm2.110.
[16] Cheng J, Zhang J, Han Y, Wang X, Ye X, Meng Y, et al. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. Cancer Res 2017;77(21):e91–100. https://doi.org/10.1158/0008-5472.CAN-17-0313.
[17] Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology 2020;72(6):2000–13. https://doi.org/10.1002/hep.31207.
[18] Skrede OJ, De Raedt S, Kleppe A, Hveem TS, Liestol K, Maddison J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 2020;395(10221):350–60. https://doi.org/10.1016/S0140-6736(19)32998-8.
[19] Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans Med Imag 2020. https://doi.org/10.1109/TMI.2020.3021387.
[20] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc Cvpr Ieee 2016:770–8. https://doi.org/10.1109/Cvpr.2016.90.
[21] Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol 2020;38(12):1346–66. https://doi.org/10.1200/jco.19.02309.
[22] Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2013. Ann Oncol 2013;24(9):2206–23. https://doi.org/10.1093/annonc/mdt303.
[23] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline focused update. J Clin Oncol 2018;36(20):2105–22. https://doi.org/10.1200/JCO.2018.77.8738.
[24] Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010;28(10):1684–91. https://doi.org/10.1200/jco.2009.24.9284.
[25] Klemm F, Joyce JA. Microenvironmental regulation of therapeutic response in cancer. Trends Cell Biol 2015;25(4):198–213. https://doi.org/10.1016/j.tcb.2014.11.006.
[26] Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol 2015;26(2):259–71. https://doi.org/10.1093/annonc/mdu450.
[27] Srinivas S, Fleuret F. Full-gradient representation for neural network visualization. Adv Neur In 2019;32.
[28] Gass P, Lux MP, Rauh C, Hein A, Bani MR, Fiessler C, et al. Prediction of pathological complete response and prognosis in patients with neoadjuvant treatment for triple-negative breast cancer. BMC Cancer 2018;18(1):1051. https://doi.org/10.1186/s12885-018-4925-1.
[29] Ruan M, Tian T, Rao J, Xu X, Yu B, Yang W, et al. Predictive value of tumor-infiltrating lymphocytes to pathological complete response in neoadjuvant treated triple-negative breast cancers. Diagn Pathol 2018;13(1):66. https://doi.org/10.1186/s13000-018-0743-7.
[30] Yang X, Rao J, Yang W, Shui R. Evaluation of the predictive and prognostic values of stromal tumor-infiltrating lymphocytes in HER2-positive breast cancers treated with neoadjuvant chemotherapy. Targeted Oncol 2018;13(6):757–67. https://doi.org/10.1007/s11523-018-0602-1.
[31] Abe N, Matsumoto H, Takamatsu R, Tamaki K, Takigami N, Uehara K, et al. Quantitative digital image analysis of tumor-infiltrating lymphocytes in HER2-positive breast cancer. Virchows Arch 2019;476(5):701–9. https://doi.org/10.1007/s00428-019-02730-6.
[32] Korde LA, Somerfield MR, Carey LA, Crews JR, Denduluri N, Hwang ES, et al. Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO guideline. J Clin Oncol 2021;39(13):1485–505. https://doi.org/10.1200/jco.20.03399.
[33] Dodington DW, Lagree A, Tabbarah S, Mohebpour M, Sadeghi-Naini A, Tran WT, et al. Analysis of tumor nuclear features using artificial intelligence to predict response to neoadjuvant chemotherapy in high-risk breast cancer patients. Breast Cancer Res Treat 2021;186(2):379–89. https://doi.org/10.1007/s10549-020-06093-4.
[34] Rye IH, Helland A, Saetersdal A, Naume B, Almendro V, Polyak K, et al. Intra-tumor heterogeneity as a predictor of therapy response in HER2 positive breast cancer. Cancer Res 2012;72. https://doi.org/10.1158/0008-5472.Sabcs12-P3-05-04.
[35] Chitalia RD, Rowland J, McDonald ES, Pantalone L, Cohen EA, Gastounioti A, et al. Imaging phenotypes of breast cancer heterogeneity in preoperative breast dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) scans predict 10-year recurrence. Clin Cancer Res 2020;26(4):862–9. https://doi.org/10.1158/1078-0432.CCR-18-4067.
[36] Huang X, Mai J, Huang Y, He L, Chen X, Wu X, et al. Radiomic nomogram for pretreatment prediction of pathologic complete response to neoadjuvant therapy in breast cancer: predictive value of staging contrast-enhanced CT. Clin Breast Cancer 2020. https://doi.org/10.1016/j.clbc.2020.12.004.
[37] Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. Eur J Nucl Med Mol Imag 2019;46(7):1468–77. https://doi.org/10.1007/s00259-019-04313-8.
[38] Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics 2019;9(5):1303–22. https://doi.org/10.7150/thno.30309.