# RESEARCH



# Early prediction of neoadjuvant therapy response in breast cancer using MRI-based neural networks: data from the ACRIN 6698 trial and a prospective Chinese cohort



Siyao Du<sup>1†</sup>, Wanfang Xie<sup>2,3†</sup>, Si Gao<sup>1</sup>, Ruimeng Zhao<sup>1</sup>, Huidong Wang<sup>7</sup>, Jie Tian<sup>2,3\*</sup>, Jiangang Liu<sup>2,3,4\*</sup>, Zhenyu Liu<sup>5,6\*</sup> and Lina Zhang<sup>1,7\*</sup>

# Abstract

**Background** Early prediction of treatment response to neoadjuvant therapy (NAT) in breast cancer patients can facilitate timely adjustment of treatment regimens. We aimed to develop and validate a MRI-based enhanced self-attention network (MESN) for predicting pathological complete response (pCR) based on longitudinal images at the early stage of NAT.

**Methods** Two imaging datasets were utilized: a subset from the ACRIN 6698 trial (dataset A, n = 227) and a prospective collection from a Chinese hospital (dataset B, n = 245). These datasets were divided into three cohorts: an ACRIN 6698 training cohort (n = 153) from dataset A, an ACRIN 6698 test cohort (n = 74) from dataset A, and an external test cohort (n = 245) from dataset B. The proposed MESN allowed for the integration of multiple timepoint features and extraction of dynamic information from longitudinal MR images before and after early-NAT. We also constructed the Pre model based on pre-NAT MRI features. Clinicopathological characteristics were added to these image-based models to create integrated models (MESN-C and Pre-C), and their performance was evaluated and compared.

**Results** The MESN-C yielded area under the receiver operating characteristic curve (AUC) values of 0.944 (95% CI: 0.906 – 0.973), 0.903 (95% CI: 0.815 – 0.965), and 0.861 (95% CI: 0.811 – 0.906) in the ACRIN 6698 training, ACRIN 6698 test and external test cohorts, respectively, which were significantly higher than those of the clinical model (AUC: 0.720 [95% CI: 0.587 – 0.842], 0.738 [95% CI: 0.669 – 0.796] for the two test cohorts, respectively; p < 0.05) and Pre-C

<sup>†</sup>Siyao Du and Wanfang Xie contributed equally to this work.

\*Correspondence: Jie Tian jie.tian@ia.ac.cn Jiangang Liu jgliu@buaa.edu.cn Zhenyu Liu zhenyu.liu@ia.ac.cn Lina Zhang Inzhang@cmu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article are shared in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

(AUC: 0.697 [95%CI: 0.554 – 0.819], 0.726 [95%CI: 0.666 – 0.797] for the two test cohorts, respectively; p < 0.05). High AUCs of the MESN-C maintained in the ACRIN 6698 standard (AUC = 0.853 [95%CI: 0.676 – 1.000]) and experimental (AUC = 0.905 [95%CI: 0.817 – 0.993]) subcohorts, and the interracial and external subcohort (AUC = 0.861 [95%CI: 0.811 – 0.906]). Moreover, the MESN-C increased the positive predictive value from 48.6 to 71.3% compared with Pre-C model, and maintained a high negative predictive value (80.4–86.7%).

**Conclusion** The MESN-C using longitudinal multiparametric MRI after a short-term therapy achieved favorable performance for predicting pCR, which could facilitate timely adjustment of treatment regimens, increasing the rates of pCR and avoiding toxic effects.

**Trial registration** Trial registration at https://www.chictr.org.cn/. Registration number: ChiCTR2000038578, registered September 24, 2020.

**Keywords** Breast cancer, Multiparametric MRI, Neoadjuvant therapy, Pathological complete response, Longitudinal radiomics, Deep learning

# Introduction

Neoadjuvant therapy (NAT) has become a standardof-care for patients with high-risk and locally advanced breast cancer [1], allowing downstaging the tumor and rapidly assessing drug susceptibility using longitudinal imaging. Achieving pathologic complete response (pCR) is associated with improved survival outcomes [2, 3]. Early prediction of pCR is crucial, as it can guide clinical decisions and optimize therapeutic strategies for individual patients.

Radiological evaluations such as the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria serve as non-invasive tools for evaluating tumor response [4]. Yet, the accuracy of these evaluations is frequently compromised by factors such as noncentric shrinkage pattern, inflammatory effects, and limitations in imaging resolution. Recent advancements have highlighted the potential of quantitative radiological imaging analysis, particularly multiparametric MRI, in predicting biological characteristics, gene expression profiles, treatment responses, and prognostic outcomes [5-8]. Radiomics-based quantitative analysis of pre-NAT MRI has shown potential for predicting pCR in breast cancer with satisfactory accuracy [5-7]. However, these studies have predominantly focused on single timepoint images, potentially missing vital information from longitudinal imaging that captures the tumor's dynamic changes over time. Several multicenter studies have enhanced the pCR predictive performance by incorporating multiple timepoint images, including mid-NAT [9] or post-NAT longitudinal analysis [10]. However, mid-NAT or post-NAT analyses imply longer NAT durations, which could prolong the use of ineffective drugs, exacerbate adverse effects and miss surgical opportunities for patients. Early-NAT accurate prediction can maximize patient benefits. The Breast Multiparametric MRI for prediction of neoadjuvant chemotherapy Response (BMMR2) challenge [11], based on ACRIN 6698 trial data, indicated the feasibility of using the early-NAT timepoint for pCR prediction. However, the models in the BMMR2 challenge did not fully exploit multiparametric MRI and longitudinal imaging adaptation network structures, leading to suboptimal performance. Furthermore, validation in heterogeneous cohorts is essential to comprehensively understand the clinical utility, key features, and network connections of longitudinal MRI in predicting pCR.

Multilayer Perceptron (MLP), as a classical neural network, has demonstrated superior performance in radiomics and deep learning studies [9, 10]. Naive MLP architectures can not capture interdependencies among temporal features. The self-attention mechanism directly captures the internal correlation within time-series data, dynamically adjusts attention weights for input features, and enhances feature representation capability through temporal context modeling [12]. Therefore, an enhanced self-attention module (ESM) was combined with the MLP for improving the performance of the longitudinal imaging model.

In this study, we aim to develop an early multiparametric MRI neural network model that integrates MLP and ESM, serving as a potential imaging tool for early determination of pCR in breast cancer. Additionally, since the early timepoint have not yet been clinically routinely evaluated, our local institution prospectively collected early MRI data to externally validate the model.

# **Materials and methods**

#### Patient cohorts

The overall design of this study is shown in Fig. 1. Two datasets were used in this study: a subset from the American College of Radiology Imaging Network (ACRIN) 6698 trial primary analysis cohort [13] (dataset A) and a prospective collection cohort at a Chinese hospital from September 2020 to September 2022 (dataset B). As shown in Fig. 2, the dataset A included



**Fig. 1** The overall design of the study. (a) Prediction of pCR after early-NAT in breast cancer can facilitate timely adjustment of therapy decision. (b) Participants in the ACRIN 6698 trial with useable MRI at both pre-NAT and early-NAT, were used to develop and internally test image-based models. An external test was conducted using a prospective cohort from a Chinese hospital. Clinicopathological characteristics were incorporated into the models for early pCR prediction. (c) The predictive model used a neural network-based quantitative analysis incorporating an enhanced self-attention module to capture dynamic information from longitudinal MRI before and after early-NAT. (d) The model's performance was evaluated using feature importance ranking and ablation experiments, including single time-point images, incomplete sequences, and the removal of the enhanced self-attention module. ROC and DCA curves were compared for these experiments

Abbreviations: pCR = pathological complete response; NAT = neoadjuvant therapy; ROC = receiver operating characteristic; DCA = decision curve analysis

227 participants from the ACRIN 6698 primary analvsis cohort with evaluable MRI at both pre-NAT and early-NAT. We set the test cohort to be consistent with that of the BMMR2 challenge [11] for a fair comparison, so the dataset A were split into a training cohort (153 of 227) and a test cohort (74 of 227). The initial Chinese cohort included 295 participants, with inclusion criteria as follows: (1) women 18 years of age or older with invasive breast cancers who were planning to undergo NAT; (2) available molecular typing information and received standard clinical treatment; (3) acceptable MRI at both pre-NAT and early-NAT. Following the application of exclusion criteria, 245 participants were available for analysis. Consequently, the three cohorts included (1) the ACRIN 6698 training cohort (n = 153); (2) the ACRIN 6698 test cohort (n = 74); (3) the external test cohort (n = 245). In the dataset A, the early-NAT timepoint was set to 3 weeks after initial treatment [13]. The early-NAT timepoint in the dataset B was set to the first three weeks, which is the first cycle of NAT, in order to correspond to the dataset A (Supplementary Fig. S1).

# NAT regimen and histopathology analysis

The ACRIN 6698 and the external test cohort underwent ultrasound-guided biopsy within 2 weeks prior to NAT to determine receptor status. The status of estrogen receptor (ER), progestone receptor (PR), HER2, and Ki-67 index were determined by immunohistochemistry (IHC) (Supplementary Material-I). The molecular subtypes were classified into three subtypes: HR+/HER2-, HER2+, HR-/HER2-, and the NAT regimen was determined according to molecular subtypes. The ACRIN 6698 trial had a designed medication dosing interval and the participants were divided into a standard subcohort or an experimental



Fig. 2 Patient enrollment flowchart

subcohort. The standard regimen was 12 cycles of weekly paclitaxel for 12 weeks, followed by four cycles of anthracycline-cyclophosphamide before the surgical procedure. The experimental regimen added one of nine experimental agents in 12 cycles of weekly paclitaxel [13]. In the external test cohort, the participants underwent standard six or eight cycles of NAT (3 weeks for each cycle). The NAT regimens were either anthracycline based, taxane based or anthracycline and taxane based according to the National Comprehensive Cancer Network (NCCN) guideline [14]. For HER2-positive tumors, the ACRIN 6698 trial received trastuzumab for the first 12 weeks, while in the external test cohort, anti-HER2 targeted trastuzumab (H) or trastuzumab + pertuzumab (HP) were added to the chemotherapy drugs throughout the entire cycle.

The status of pCR for each target tumor was determined by surgical-pathological results within 1 month after NAT. pCR was defined as the absence of residual invasive tumor in both the breast and axillary lymph nodes, while residual ductal carcinoma in situ was allowed (ypT0/Tis ypN0) [13, 15].

#### Image acquisition and tumor segmentation

The peak phase of DCE-MRI (hereinafter referred to as T1) and DWI (hereinafter referred to as DW) of the ACRIN 6698 and external cohorts were used for image analysis. Acquisition parameters are shown in Supplementary Material-II and Table S1. The target tumor of each patient was visible in both pre-NAT and early-NAT MRI. Tumor volume masks on T1 images were created by threshold-based, seed point driven, semi-automatic segmentation using ITK-SNAP (www. itksnap.org) [16]. Due to the lower resolution of DW compared to T1, radiologists manually draw tumor borders on each slice of DW ( $b = 800 \text{ s/mm}^2$ ). Tumor borders of DW referenced to the apparent diffusion coefficient (ADC) map to ensure clear demarcation between diffusion-restricted and normal areas, then they shared the ROI. To better assess tumor heterogeneity, the ROIs of T1, DW, and ADC encompassed the entire tumor including necrotic, hemorrhagic areas and surrounding spiculation [17-19]. Two radiologists with at least five years' experience in breast imaging independently segmented tumor with T1, DW and ADC. If multicentric lesions located in different quadrants, the largest one was selected as the object. If multifocal lesions located in the same quadrant, all lesions were included in the ROI mask.

# Imaging feature extraction and selection

Before imaging features extraction, the original MRI and mask images were processed by the B-spline interpolation method to normalize the spatial resolution to  $1 \times 1 \times 1$  mm<sup>3</sup> [4]. The PyRadiomics package (v3.1.0) [20] was utilized to extract imaging features,

including intensity-based histogram, shape, and various gray level metrics from pre-NAT and early-NAT MRI (Supplementary Material-III). Recent studies [7, 21] reported that combining the intratumoral and peritumoral information could enhance model performance. Thus, a 5 mm distance [7, 22] of peritumoral area was expanded using the morphological dilation operation (a function from the Scipy package v1.8.0) for the extraction of peritumoral features. A total of 1409 features were extracted per ROI, resulting in 16,908 features per patient from the intratumoral and peritumoral areas of the pre-NAT and early-NAT T1, DW and ADC, respectively.

The imaging features extracted from the pre-NAT and early-NAT MRI were defined as the pre-features and early-features, respectively. We calculated the delta-features which were the differences between the pre-features and the early-features of corresponding sequence.

$$Delta - features = (early - features) - (pre - features)$$
(1)

Following feature extraction, repeatability analysis was performed on a subset of 30 samples randomly chosen from the external test cohort using intraclass correlation coefficient (ICC). Features with an ICC value > 0.80 [23] were considered to have satisfactory reproducibility and were retained for further analysis. Also, spearman correlation analysis eliminated highly correlated feature with correlation coefficients > 0.80. The remaining features were normalized using Z-score normalization.

# Development of MRI-based enhanced self-attention network (MESN)

An MRI-based enhanced self-attention network (MESN) was developed to predict pCR using multiparameter MRI at the early stage of NAT. The flowchart of the construction of the MESN was shown in Fig. 3. For each patient, the pre-features (i.e.,  $Pre_{T1}$ ,  $Pre_{DW}$ , and  $Pre_{ADC}$ ), the early-features (i.e.,  $Early_{T1}$ ,  $Early_{DW}$ ,  $Early_{ADC}$ ), and the delta-features (i.e.,  $Delta_{T1}$ ,  $Delta_{DW}$ ,  $Delta_{ADC}$ ) were fed into the MESN, and the output of the MESN was the pCR prediction probability (denoted as the signature of image-based model).

The main components in the MESN were the nine multi-layer perceptions (MLP1.T1, MLP1.DW, MLP1. ADC, MLP2.T1, MLP2.DW, MLP2.ADC, MLP3.T1, MLP3.DW, and MLP3.ADC) and a proposed ESM. The MLPs were implemented to simulate the nonlinear relationship between the input features and output probability. For the ESM, the positional encoding module was used to encode the positional information of the  $Pre_{all}$ ,  $Early_{all}$ , and  $Delta_{all}$  features, and the multi-head self-attention module [12] from the selfattention-cv package (v 1.2.3) [24] was utilized to learn the temporal relationship among the  $Pre_{all}$ ,  $Early_{all}$ , and  $Delta_{all}$  features. Additionally, two loss functions were used in the ESM, one was the cross-entropy loss which was utilized to train the model to discriminate pCR samples from non-pCR samples, the other was the proposed feature independence loss which was utilized to regularize the  $Pre_{all}$ ,  $Early_{all}$  and  $Delta_{all}$  features to be discriminated for the model



Fig. 3 The MRI-based enhanced self-attention network model for pCR prediction

Abbreviations: pCR = pathological complete response; NAT = neoadjuvant therapy; MLP = multi-layer perceptions; T1 = T1-weighted dynamic contrast-enhanced MRI; DW = diffusion weighted imaging; ADC = apparent diffusion coefficient; ATP = All Time Points

optimization. The details of the MESN can be seen in Supplementary Material-IV.

#### Ablation analysis for MESN

To explore whether different components in the proposed MESN were required to accurately predict pCR, we conducted a set of ablation experiments by modifying the input image information and network structure (Supplementary Material-V and Table S2). First, we only retained the multiparametric MRI features at a single timepoint (pre-features or early-features) or delta-features as inputs to evaluate the model performance (denoted as Pre, Early and Delta). Then, we separately input individual T1, T1 + DWI, or T1 + ADC sequence into MESN to evaluate the performance with different MRI sequences (denoted as MESN [T1], MESN [T1 + DW] and MESN [T1 + ADC]). Finally, we removed ESM to evaluate the pCR prediction performance (denoted as NoESM).

## Integration with clinicopathological characteristics

Univariable analysis was performed to identify clinicopathological characteristics associated with pCR. Multivariable logistic regression analysis was conducted to construct a clinical model. Then, we further developed integration models with the signature of image-based models and clinicopathological characteristics (Supplementary Material-VI and Table S3-S4), denoted as MESN-C, Pre-C, Early-C, Delta-C, MESN-C (T1), MESN-C (T1+DW), MESN-C (T1+ADC), and NoESM-C.

# **Clinical benefit analysis**

At pre-NAT, only the Pre and Pre-C model were available. MESN-C was available after early-NAT. To clarify the clinical benefit of MESN-C, we compared the benefit proportion between the optimal pre-NAT models and the MESN-C. For patients predicted by the models as pCR or non-pCR, we used the actual pathological results as the gold standard to assess how many patients could benefit from MESN-C.

# Statistical analysis

Clinicopathological characteristics and pre-NAT MRI findings were compared using Mann–Whitney U test and chi-square (or Fisher's exact) test. In the two test cohorts, the model performance was assessed using the area under the receiver operating characteristic (ROC) curve, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The delong test was used to compare the differences between the ROC curves [25]. The net reclassification index (NRI) and integrated discrimination improvement (IDI) were used to evaluate improvements in the predictive models. The accuracy, sensitivity, specificity, PPV and NPV were calculated at a threshold of 0.50. The model's clinical utility was determined using decision curve analysis (DCA). All statistical analyses were performed using Python 3.8.18. A *p*-value < 0.05 indicated statistical significance.

### Results

# **Baseline characteristics**

A total of 472 patients with 944 MRI examinations were finally analyzed in this study. After NAT, 48/153 (31.4%) in the ACRIN 6698 training cohort, 23/74 (31.1%) in the ACRIN 6698 test cohort and 79/245 (32.2%) in the external test cohort achieved pCR, which showed no significant difference between them (Table 1, p = 0.965 and 0.856). The external test cohort had a higher proportion of grade III tumors (p < 0.001) and HER2-positive subtypes (p = 0.002)compared with the ACRIN 6698 training cohort. The patient characteristics between pCR and non-pCR are summarized in Supplementary Table S5. pCR tended to present grade III tumor and high Ki-67 only in the external test cohort (p < 0.01). HR, HER2, molecular subtype showed significant associations with pCR in both of the ACRIN 6698 and external test cohorts. The ACRIN 6698 data revealed that HR and HER2 were key clinicopathological characteristics for pCR prediction to build the clinical model, which obtained the AUC of 0.713 (95% confidence interval [CI]: 0.604 - 0.744) in the ACRIN 6698 training, 0.720 (95% CI: 0.587-0.842) in the ACRIN 6698 test and 0.738 (95% CI: 0.669-0.796) in the external test cohort (Table 2).

#### Imaging feature selection

We removed the features with poor repeatability and high correlation from the extracted features, resulting in 198 pre-features of T1, 203 pre-features of DW, 207 pre-features of ADC; 204 early-features of T1, 200 early-features of DW, 216 early-features of ADC; 321 delta-features of T1, 322 delta-features of DW, 334 delta-features of ADC. All the above features were used as inputs for MESN construction. The permutation analysis was used to evaluate the feature importance (Fig. 4a and Supplementary Fig. S2), with selected top three features illustrated in Fig. 4b-g. The top 30 important features mainly came from early timepoint (23/30), T1 sequence (19/30), and peritumoral region (17/30).

#### Performance of the MESN and MESN-C

The proposed MESN model obtained a high accuracy with an AUC of 0.923 (95%CI: 0.874–0.963) in the ACRIN 6698 training cohort. Furthermore, the MESN

Characteristics	ACRIN 6698 Training (n = 153)	ACRIN 6698 Test (n = 74)	External Test (n = 245)	<b>p</b> *	p#
Age (years)	48±15	48±14	51±15	0.912	0.062
Tumor size (cm)	$3.40 \pm 1.90$	$3.85 \pm 2.70$	$3.70 \pm 2.40$	0.018	0.298
Race				0.807	NA
White	109 (71.2)	53 (71.6)	0		
Black	16 (10.5)	9 (12.2)	0		
Asian	11 (7.2)	3 (4.0)	245 (100)		
Other	17 (11.1)	9 (12.2)	0		
TNM stage				NA	NA
II A	0	0	30 (12.2)		
II B	0	0	95 (38.8)		
III A	0	0	50 (20.4)		
III B	0	0	29 (11.8)		
III C	0	0	41 (16.8)		
Not available	153 (100)	74 (100)	0		
Nuclear grades				0.648	< 0.001
1	3 (2.0)	2 (2.7)	1 (0.4)		
2	45 (29.4)	17 (23.0)	156 (63.7)		
3	104 (68.0)	55 (74.3)	88 (35.9)		
Missing	1 (0.6)	0	0		
Enhancement type				0.101	0.053
Mass	141 (92.2)	63 (85.1)	213 (86.9)		
Non-mass	12 (7.8)	11 (14.9)	32 (13.1)		
Multi-focality or multi-centricity				0.494	0.269
Yes	88 (57.5)	39 (52.7)	127 (51.8)		
No	65 (42.5)	35 (47.3)	118 (48.2)		
HR				0.448	0.306
Positive	87 (56.9)	46 (62.2)	152 (62.0)		
Negative	66 (43.1)	28 (37.8)	93 (38.0)		
HER2				0.769	0.002
Positive	40 (26.1)	18 (24.3)	102 (41.6)		
Negative	113 (73.9)	56 (75.7)	143 (58.4)		
Ki-67				NA	NA
≤20%	0	0	50 (20.4)		
>20%	0	0	195 (79.6)		
Not available	153 (100)	74 (100)	0		
Molecular subtypes				0.300	0.003
Luminal HER2-	62 (40.5)	33 (44.6)	95 (38.8)		
Luminal HER2+	25 (16.3)	13 (17.6)	57 (23.3)		
HER2 enriched	15 (9.8)	5 (6.7)	45 (18.4)		
TNBC	51 (33.4)	23 (31.1)	48 (19.5)		
Treatment response				0.965	0.856
pCR	48 (31.4)	23 (31.1)	79 (32.2)		
non-pCR	105 (68.6)	51 (68.9)	166 (67.8)		

Table 1 Comparison of clinicopathologic characteristics among the ACRIN 6698 training, ACRIN 6698 test and the external test cohorts

Note. Age and tumor size are shown as median  $\pm$  IQR, and the others are shown as proportions (percentages).  $p^*$  values represent the comparison of clinicopathologic variables between the ACRIN 6698 training and test cohort;  $p^{\#}$  values represent the comparison of clinicopathologic variables between the ACRIN 6698 training and test cohort;  $p^{\#}$  values represent the comparison of clinicopathologic variables between the ACRIN 6698 training and test cohort. Bold text with p values indicate statistical significance. NA = not applicable; ER = estrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor-2; TNBC = triple-negative breast cancer

achieved consistently a high accuracy with AUC values of 0.860 (95%CI: 0.757 - 0.939) in the ACRIN 6698 test cohort and 0.804 (95%CI: 0.752 - 0.864) in the external test cohort (Table 2). After incorporating clinicopathological characteristics, the MESN-C

performed remarkably well with AUC of 0.903 (95%CI: 0.815 - 0.965) and 0.861 (95%CI: 0.811 - 0.906) in the two test cohorts, respectively.

In the subgroup analysis, MESN-C maintained high predictive performance in the subgroups of different T

Table 2     The model performance for pCR prediction in	n the ACRIN 6698 training,	, internal ACRIN 6698 test a	nd external test cohort with
full sequences under different early-treatment timepo	ints		

Models	AUC (95% CI)	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
ACRIN 6698 Training						
Clinical	0.713 (0.604, 0.744)	61.4	83.3	51.4	44.0	87.1
Pre	0.873 (0.811, 0.925)	79.7	77.1	81.0	64.9	88.5
Pre-C	0.896 (0.841, 0.943)	80.4	81.2	80.0	65.0	90.3
MESN	0.923 (0.874, 0.963)	86.3	77.1	90.5	78.7	89.6
MESN-C	0.944 (0.906, 0.973)	83.7	83.3	83.8	70.2	91.7
ACRIN 6698 Test						
Clinical	0.720 (0.587, 0.842)	62.2	78.3	54.9	43.9	84.8
Pre	0.626 (0.475, 0.754)	64.9	34.8	78.4	42.1	72.7
Pre-C	0.697 (0.554, 0.819)	75.7	65.2	80.4	60.0	83.7
MESN	0.860 (0.757, 0.939)	82.4	56.5	94.1	81.2	82.8
MESN-C	0.903 (0.815, 0.965)	86.5	73.9	92.2	81.0	88.7
External Test						
Clinical	0.738 (0.669, 0.796)	61.2	84.8	50.0	44.7	87.4
Pre	0.607 (0.542, 0.685)	60.0	45.6	66.9	39.6	72.1
Pre-C	0.726 (0.666, 0.797)	66.5	65.8	66.9	48.6	80.4
MESN	0.804 (0.752, 0.864)	77.6	46.8	92.2	74.0	78.5
MESN-C	0.861 (0.811, 0.906)	81.6	72.2	86.1	71.3	86.7

Note. Numbers in parentheses are 95% confidence intervals; Clinical: model based on clinicopathological characteristics; Pre: model based on pre-features; Pre-C: model built using the signature of Pre and clinicopathological characteristics; MESN: MRI-based enhanced self-attention network; MESN-C: model built using the signature of MESN and clinicopathological characteristics

pCR=pathological complete response; AUC=receiver operating characteristic curve; ACC=accuracy; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value

stages, nuclear grades, HR, HER2 status and molecular subtypes (Supplementary Table S6). Notably, MESN-C was not affected by NAT regimens, as demonstrated by the performance of the ACRIN 6698 standard (AUC = 0.853, 95%CI: 0.676 - 1.000), experimental (AUC = 0.905, 95%CI: 0.817 - 0.993), and the external subcohorts (AUC = 0.861, 95%CI: 0.811 - 0.906).

We also evaluated the clinical utility of MESN-C in pCR prediction (Supplementary Fig. S3), which indicated that the MESN-C was clinically useful when intervention was decided in the threshold range of 0-85% in the ACRIN 6698 and external test cohorts.

#### Ablations of MESN and model comparisons

The MESN model, integrating multiparametric features from longitudinal MR images, showed a superior performance in predicting pCR compared with the conventional clinical model and single timepoint models (Tables 2 and 3, Supplementary Table S7 and Fig. 5).

Compared with the clinical model, the predictive performance of MESN-C was the highest in the ACRIN 6698 test cohort (delong/NRI/IDI p = 0.007/0.027/<0.001) and the external test cohort (delong/NRI/IDI p < 0.001, Table 3).

To illustrate the advantages of using full timepoint in the prediction of pCR, MESN was compared with the Pre, Early and Delta models. Unsurprisingly, MESN achieved a better performance than Pre / Early / Delta models (AUC = 0.626 [95%CI: 0.475 - 0.754] / 0.725 [95%CI: 0.604 - 0.850] / 0.640 [95%CI: 0.489 - 0.779] in the ACRIN 6698 test, respectively; AUC = 0.607[95%CI: 0.542 - 0.685] / 0.693 [95%CI: 0.632 - 0.764] / 0.682 [95%CI: 0.603 – 0.752] in the external test cohort, respectively). The MESN-C also achieved a better predictive performance than Pre-C / Early-C / Delta-C models (AUC = 0.697 [95%CI: 0.554 - 0.819] / 0.796 [95%CI: 0.681 - 0.890] / 0.680 [95%CI: 0.527 - 0.810] in the ACRIN 6698 test,, respectively; AUC = 0.726[95%CI: 0.666 - 0.797] / 0.822 [95%CI: 0.768 - 0.875] / 0.750 [95%CI: 0.675-0.812] in the external test cohort,, respectively) (Tables 2 and 3, Supplementary Table S7 and Fig. 5). The DCA of MESN and MESN-C maintained the highest net benefit within a wider range of risk threshold (Supplementary Fig. S3).

To investigate the necessity of using full sequences in the prediction of pCR, MESN was compared with MESN(T1), MESN(T1 + DW) and MESN(T1 + ADC) models. There was a significant improvement in prediction performance of MESN (Table 3). MESN(T1 + ADC) or MESN-C(T1 + ADC) has the closest AUC and DCA to MESN or MESN-C, respectively (Supplementary Table S7, Fig. 5 and Supplementary Fig. S3), indicating T1 and ADC sequences are relatively important. Feature importance analysis also



Fig. 4 (See legend on next page.)

#### (See figure on previous page.)

**Fig. 4** (a) The importance ranking of the top 30 features in the MESN. (**b**-**g**) The original T1, DW, ADC and corresponding top three radiomics feature maps for patients with pCR (**b**, **d**, **f**) and non-pCR (**c**, **e**, **g**) who had the same tumor stage, NAT regimen and the similar age. The differences of raw data were not apparent to the naked eye. The MESN model recognized the subtle differences: peritumoral logarithm\_ngtdm\_Contrast and wavelet-LLL\_glszm\_Low-GrayLevelZoneEmphasis on T1 images after early-NAT in pCR patients is lower than in non-pCR patients, whereas logarithm\_firstorder\_Minimum is higher. The color bar on the right refers to the image intensity value of the radiomics feature map, with red representing high values and blue representing low values

Abbreviations: MESN = MRI-based enhanced self-attention network; pCR = pathological complete response; NAT = neoadjuvant therapy; T1 = T1-weighted dynamic contrast-enhanced MRI; <math>DW = diffusion weighted imaging; ADC = apparent diffusion coefficient; HR = hormone receptor; HER2 = human epidermal growth factor receptor-2

indicated that the top 30 features all came from T1 or ADC (Fig. 4a).

To estimate the importance of capturing the dynamic information from longitudinal MR images, we compared MESN with NoESM. MESN performed better (AUC = 0.860 [95%CI: 0.757 - 0.939] vs. 0.671 [95%CI: 0.511 - 0.800] in the ACRIN 6698 test; AUC = 0.804 [95%CI: 0.752 - 0.864] vs. 0.754 [95%CI: 0.698 - 0.823] in the external test cohort) (Supplementary Table S7, Fig. 5) and showed higher clinical net benefits than NoESM (Supplementary Fig. S3). After integrating clinicopathological characteristics, the superiority of MESN-C remained significant (Table 3and Fig. 5).

Additionally, to surpass the BMMR2 challenge, we compared our models with the models in the BMMR2 challenge using the same ACRIN 6698 test settings. The AUC of the MESN-C (AUC = 0.903 [95%CI: 0.815 - 0.965]) model was higher than those of the top three teams in the BMMR2 challenge (team A / B / C, AUC = 0.840 [95%CI: 0.748 - 0.932] / 0.838 [95%CI: 0.748 - 0.928] / 0.803 [95%CI: 0.702 - 0.904]) [11]. The comparison among the models was shown in Table 4.

# **Clinical benefit analysis**

As illustrated in Fig. 6, for the total 245 patients in the external test cohort, Pre-C predicted 107 patients as pCR and correctly identified 48.6% (52/107) true-positive cases. MESN-C increased the PPV to 71.3% (57/80). MESN-C also slightly improved the NPV from 80.4% (111/138) to 86.7% (143/165). These patients could benefit from MESN-C to decide whether to continue with the initial medication or adjust therapy to achieve pCR.

# Discussion

In this study, longitudinal MR images were used from the ACRIN 6698 trial data to construct a neural network model (MESN) for early pCR prediction in breast cancer, and this model performed well in both the ACRIN 6698 test cohort and the external test cohort. After integrating the signature of MESN with the clinicopathological characteristics, the MESN-C model achieved satisfactory performance in AUC, sensitivity, and specificity, offering clinicians the potential to promptly adjust treatment, thereby enhancing pCR rates and mitigating the risk of adverse effects from unnecessary treatments.

Changes in tumor size, cellularity, and perfusion following treatment are indicative of the tumor's response to therapy [26], and these alterations can be effectively monitored through longitudinal multiparametric MRI [9, 10, 13, 27-29]. Consequently, incorporating early multiparametric MRI could facilitate the prediction of pCR. MESN was developed by integrating pre-NAT and early-NAT multiparametric MRI, capturing the dynamic changes therein. MESN demonstrated a superior prediction performance compared to that of the models with single timepoint data, partial sequence inputs, or those lacking the ESM module. Huang et al. [10] have previously assessed treatment responses using deep learning based on pre- and post-NAT MRI, achieving AUC ranging from 0.837 to 0.929 across breast cancer subtypes. Li et al. [9] setting the predictive timepoint at mid-NAT, reported an AUC exceeding 0.90 with their proposed artificial intelligence system. These two models required continuation until the mid or late stages of NAT before becoming operational. This extended observation period could potentially prolong the administration of ineffective medications and heighten the risk of side effects. Early MRI assessments, while not mandatory, necessitate prospective data collection, similar to the one-cycle data utilized in the external test cohort. The ACRIN 6698 trial, a publicly available dataset rich in early longitudinal multiparametric MRI, served as the foundation for the BMMR2 challenge's pCR prediction competition [11]. MESN-C outperformed the best model in the BMMR2 challenge in the same ACRIN 6698 test set, with an AUC of 0.903 (95%CI: 0.815-0.965) compared to that of 0.840 (95%CI: 0.748 - 0.932), a performance difference that could be attributed to the critical inclusion of multiple timepoints, multiparametric MRI inputs, and the integration of an ESM module. An additional benefit of the MESN-C is its lack of reliance on additional kinetic mapping calculations, which are often challenging to standardize across different settings. Moreover, the expanded external testing is deemed a crucial milestone before these models can be clinically integrated to inform therapeutic decisions [11]. Our prospective

Table 3 Model comparisons in term of AUC, net reclassification index (NRI) and integrated discrimination improvement (IDI) in the
internal ACRIN 6698 test cohort and the external test cohort

Model comparisons	Delong	NR	I	IDI	
	p	95%Cl	р	95%Cl	р
ACRIN 6698 test					
MESN-C vs. Clinical	0.007	0.329 (0.036-0.622)	0.027	0.298 (0.178-0.418)	< 0.001
MESN vs. Pre	0.001	0.374 (0.083-0.666)	0.012	0.119 (0.038-0.200)	0.004
MESN-C vs. Pre-C	0.001	0.205 (-0.055–0.465)	0.123	0.255 (0.120-0.390)	< 0.001
MESN vs. MESN (T1)	0.006	0.296 (0.063–0.529)	0.013	0.117 (0.070-0.164)	< 0.001
MESN-C vs. MESN-C (T1)	0.014	0.283 (0.044-0.522)	0.020	0.167 (0.046–0.289)	0.007
MESN vs. MESN (T1 + DW)	0.004	0.311 (0.080-0.542)	0.008	0.076 (0.029–0.122)	0.002
MESN-C vs. MESN-C (T1 + DW)	0.017	0.122 (-0.051–0.295)	0.167	0.161 (0.062–0.259)	0.001
MESN vs. MESN (T1 + ADC)	0.704	0.013 (-0.210-0.238)	0.905	0.052 (0.017-0.088)	0.004
MESN-C vs. MESN-C (T1 + ADC)	0.253	0.228 (0.014-0.443)	0.037	0.162 (0.077-0.246)	< 0.001
MESN vs. NoESM	0.004	0.387 (0.171-0.603)	< 0.001	0.089 (0.033-0.146)	0.002
MESN-C vs. NoESM-C	0.007	0.280 (0.048-0.513)	0.018	0.238 (0.123-0.346)	< 0.001
External test					
MESN-C vs. Clinical	< 0.001	0.235 (0.111-0.358)	< 0.001	0.224 (0.164-0.284)	< 0.001
MESN vs. Pre	< 0.001	0.266 (0.110-0.421)	0.001	0.086 (0.053-0.119)	< 0.001
MESN-C vs. Pre-C	< 0.001	0.256 (0.112-0.401)	0.001	0.194 (0.124–0.264)	< 0.001
MESN vs. MESN (T1)	0.003	0.101 (-0.018-0.222)	0.096	0.068 (0.043-0.094)	< 0.001
MESN-C vs. MESN-C (T1)	0.024	0.121 (0.015-0.228)	0.026	0.086 (0.032-0.140)	0.002
MESN vs. MESN (T1 + DW)	0.007	0.051 (-0.065–0.168)	0.388	0.036 (0.012-0.061)	0.004
MESN-C vs. MESN-C (T1 + DW)	0.030	0.100 (-0.008-0.209)	0.071	0.057 (0.008–0.106)	0.022
MESN vs. MESN (T1 + ADC)	0.057	0.004 (-0.101-0.110)	0.939	0.056 (0.040-0.073)	< 0.001
MESN-C vs. MESN-C (T1 + ADC)	0.212	0.017 (-0.054-0.089)	0.633	0.120 (0.084–0.157)	< 0.001
MESN vs. NoESM	0.058	0.140 (0.035-0.245)	0.009	0.025 (0.004-0.046)	0.020
MESN-C vs. NoESM-C	0.058	0.061 (-0.035–0.156)	0.211	0.112 (0.063–0.161)	< 0.001

Note. Significant *p*-values (p < 0.05) are highlighted in bold. Clinical: model based on clinicopathological characteristics; Pre: model built based on pre-features; Pre-C: model built using the signature of Pre and clinicopathological characteristics; MESN: MRI-based enhanced self-attention network; MESN-C: model built using the signature of MESN and clinicopathological characteristics. NoESM: MESN model without enhanced self-attention module (ESM); NoESM-C: model built using the signature of NoESM and clinicopathological characteristics. The parentheses after MESN or MESN-C contain the MRI sequences input into the model

T1 = the peak phase of T1-weighted dynamic contrast enhanced MRI; DW = diffusion weighted imaging; ADC = apparent diffusion coefficient

data has proved the significant clinical advantages of the MESN-C, highlighting its potential to revolutionize patient care.

The MESN-C model, constructed by data from the ACRIN 6698 trial, demonstrated satisfactory performance in both standard and experimental subcohorts, as well as in an interracial Chinese cohort treated with NCCN-standard regimens. This indicates that the MESN-C captures early drug efficacy information that is not confined to specific NAT regimens. The MESN-C model outperformed the Pre-C model, achieving an AUC of 0.861 (95%CI: 0.811-0.906) in the interracial Chinese cohort. Compared to the Pre-C model, the MESN-C model helps identify more patients who will benefit most from NAT at the earliest stages of treatment, ensuring that these patients complete the full NAT cycles and achieve the best effects, potentially benefiting from breast-conserving surgery and omission of axillary node dissection. The MESN-C model can also identify patients with poor response after early treatment, allowing for timely adjustments of treatment protocols or alterations of therapeutic targets, thereby avoiding longer exposure to ineffective treatments, which could potentially increase the overall pCR rate and reduce the risk of recurrence. Also, imaging models that involve experimental drugs have the potential to function as early biomarkers in clinical trials, thereby expediting the validation of new drug efficacy in future research.

An interesting finding is that the top contributing features of MESN were derived from T1 images after early treatment. As previously reported, singletimepoint models using T1 images after early- [28, 30], mid- [9], and post-NAT [10] generally performed better than pre-NAT model and delta-models, or had the higher weights in longitudinal integrated models. In sequence comparisons, Eun et al. [19] reported that radiomic models built with T1 sequence outperformed those built with T2WI/DWI/ADC sequences. The importance of T1-derived features after early-NAT stems from their direct origin and the incorporation of therapeutic information. These features may potentially reflect a series of treatment-induced changes, such as microvascular remodeling [31], oxygen supply



Fig. 5 Performances of MESN and MESN-C for pCR prediction. (a-b) MESN vs. models with single timepoint input; (c-d) MESN-C vs. models with single timepoint input; (e-f) MESN vs. models with single or dual sequence input; (g-h) MESN-C vs. models with single or dual sequence input; (i-j) MESN vs. model without ESM; (k-l) MESN-C vs. model without ESM

Abbreviations: Pre: model based on pre-features; Pre-C: model built using the signature of Pre and clinicopathological characteristics; Delta: model based on delta-features; Delta-C: model built using the signature of Delta and clinicopathological characteristics; Early: model based on early-features; Early-C: model built using the signature of Early and clinicopathological characteristics; MESN: MRI-based enhanced self-attention network; MESN-C: model built using the signature of MESN and clinicopathological characteristics; The parentheses after MESN or MESN-C contain the MRI sequences input into the model (T1 = T1-weighted dynamic contrast-enhanced MRI; DW = diffusion weighted imaging; ADC = apparent diffusion coefficient); NoESM: MESN model without enhanced self-attention module (ESM); NoESM-C: model built using the signature of NoESM and clinicopathological characteristics; AUC = area under the curve

|--|

	Prediction Model	Sequence Used	Timepoint Used	Clinical Feature	Was Model Tested Using Other Data?	Model Performance
Team A	Logistic regression	SER map and kinetic maps from DCE image	Pre-NAT and early-NAT	HR and HER2	No	ACRIN 6698 test: 0.840 (95%Cl: 0.748–0.932)
Team B	XGBoost, Random forest, Logistic regression	DCE image, SER map, PE map, DW image	Pre-NAT, early-NAT, and mid-NAT	Age, Race, Le- sion Type, SBR Tumor Grade, MRI Longest Diameter, HR and HER2	No	ACRIN 6698 test: 0.838 (95%CI: 0.748–0,928)
Team C	Logistic regression	DW image	early-NAT,	MRI Longest Diameter, HR and HER2	No	ACRIN 6698 test: 0.803 (95%CI: 0.702–0.904)
MESN	Neural networks	DCE image, DW image, and ADC image	Pre-NAT and early-NAT	None	Yes	ACRIN 6698 test: 0.860 (95%Cl: 0.757–0.939) External test: 0.804 (95%Cl: 0.752–0.864)
MESN-C	Neural networks, Logistic regression	DCE image, DW image, and ADC image	Pre-NAT and early-NAT	HR and HER2	Yes	ACRIN 6698 test: 0.903 (95%Cl: 0.815– 0.965) External Test: 0.861 (95%Cl: 0.811–0.906)

MESN: MRI-based enhanced self-attention network; MESN-C: model built using the signature of MESN and clinicopathological characteristics DCE=dynamic contrast-enhanced, SER=signal enhancement ratio; DW= diffusion-weighted, HR=hormone receptor, HER2=human epidermal growth receptor 2



Fig. 6 Clinical benefit assessment of MESN-C. (a) Recommendation for pCR prediction according to Pre-C and MESN-C; (b) Benefit rate of non-pCR patients predicted in two models; (c) Benefit rate of pCR patients predicted in two models

Abbreviations: Pre: model based on pre-features; Pre-C: model built using the signature of Pre and clinicopathological characteristics; MESN: MRI-based enhanced self-attention network; MESN-C: model built using the signature of MESN and clinicopathological characteristics; pCR = pathological complete response; NAT = neoadjuvant therapy

[32], and immune infiltration [17, 21]. Histopathological analysis has demonstrated that early on-treatment immune response is more predictive of treatment outcome than baseline immune response [26, 33]. The pairing and integration of early on-treatment imaging and histopathological data are anticipated to elucidate these high-contributing features. Additionally, we also found that the peritumoral early-features of T1 images remained in the top four. The peritumoral area on T1 images better predicted pCR than the tumor area alone and associated with the immune microenvironment [7, 21, 34]. Here we newly identified a series of peritumoral features from early-NAT T1 images that also contribute to pCR prediction. The peritumoral area after early-NAT is the key region in the tumor bed after initial tumor regression. Peritumoral earlyfeatures may potentially capture the microstructure change, such as tumor cell activity, stromal fibrosis, and immune response [35].

Our study has several limitations. First, the sample size of the training cohort is limited. Future researches should aim to collect a larger number of samples from multiple centers to facilitate more robust analyses, such as the development of three-dimensional convolutional networks and the evaluation of various treatment regimens. Second, the external cohort did not include the use of experimental drugs and differed from the ACRIN 6698 trial in terms of medication frequency. Future assessments of early prediction models for specific medications are warranted as these drugs become integrated into clinical neoadjuvant settings, enabling the evaluation of larger datasets. Finally, the biological explanations directly related to MRI phenotypes have not been studied. Further studies require an interdisciplinary approach that combines genomics or pathomics to elucidate the biological meaning of phenotypes [36, 37].

# Conclusions

The proposed MESN-C using longitudinal multiparametric MRI achieved favorable performance for predicting pCR after short-term testing therapy. The MESN-C had the potential to assist clinicians in the early adjustment of therapy, increasing the rates of pCR and avoiding toxic effects.

#### Abbreviations

- MESN MRI-based enhanced self-attention network
- T1 T1-weighted dynamic contrast-enhanced MRI
- DW Diffusion weighted imaging
- ADC Apparent diffusion coefficient
- NAT Neoadjuvant therapy
- pCR Pathological complete response
- ROC Receiver operating characteristic

# **Supplementary Information**

The online version contains supplementary material available at https://doi.or g/10.1186/s13058-025-02009-6.

Supplementary Material 1

#### Acknowledgements

The authors thank Min Zhao and Lizhi Xie from GE Healthcare for the technical support.

#### Author contributions

SD and WX were involved in investigation, methodology, project administration, validation, visualization, and writing—original draft. SG, RZ and HW was involved in data curation, methodology, project administration, and writing—review & editing. ZL and JL was involved in validation, visualization, methodology, and writing—review & editing. JT was involved in investigation, supervision, and writing—review & editing. LZ was involved in conceptualization, data curation, supervision, project administration, funding acquisition, and writing—review & editing. All authors have read and agreed to the published version of the manuscript.

#### Funding

This study has received funding from National Natural Science Foundation of China (82371947, 62333022, 62027901, 82302165), the Nature Science Foundation of Beijing (Z200027).

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

#### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the First Hospital of China Medical University (Ethic code: 2019-33-2 with date of approval 6 March 2019). Participants were enrolled after providing their written informed consent.

#### **Consent for publication**

All authors agreed with the content of the present paper and consent to submit.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Radiology, The First Hospital of China Medical University, Shenyang 110001, Liaoning Province, China  <sup>2</sup>School of Engineering Medicine, Beihang University, Beijing 100191, People's Republic of China
<sup>3</sup>Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology of the People's Republic of China, Beijing 100191, People's Republic of China
<sup>4</sup>Beijing Engineering Research Center of Cardiovascular Wisdom Diagnosis and Treatment, Beijing 100029, People's Republic of China
<sup>5</sup>CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing 100190, People's Republic of China
<sup>6</sup>University of Chines Academy of Sciences, Beijing 100080, People's Republic of China
<sup>7</sup>Department of Radiology, The Fourth Affiliated Hospital of China Medical

University, Shenyang 110165, Liaoning Province, China

## Received: 8 December 2024 / Accepted: 24 March 2025 Published online: 03 April 2025

#### References

- Gralow JR, Burstein HJ, Wood W, Hortobagyi GN, Gianni L, von Minckwitz G, Buzdar AU, Smith IE, Symmans WF, Singh B, et al. Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease. J Clin Oncology: Official J Am Soc Clin Oncol. 2008;26(5):814–9.
- Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet (London England). 2014;384(9938):164–72.
- Symmans WF, Wei C, Gould R, Yu X, Zhang Y, Liu M, Walls A, Bousamra A, Ramineni M, Sinn B, et al. Long-Term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype. J Clin Oncology: Official J Am Soc Clin Oncol. 2017;35(10):1049–60.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J cancer (Oxford England: 1990). 2009;45(2):228–47.
- Shi Z, Huang X, Cheng Z, Xu Z, Lin H, Liu C, et al. MRI-based quantification of intratumoral heterogeneity for predicting treatment response to neoadjuvant chemotherapy in breast cancer. Radiology. 2023;308(1):e222830.
- 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. Clinical Cancer Research 2019;25(12):3538–47.
- Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, Plecha D, Madabhushi A. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast cancer Research: BCR. 2017;19(1):57.
- Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, Gu Y, Li A, Lu N, He Z, et al. Development and validation of a preoperative magnetic resonance imaging Radiomics-Based signature to predict axillary lymph node metastasis and Disease-Free survival in patients with Early-Stage breast cancer. JAMA Netw Open. 2020;3(12):e2028086.
- Li W, Huang YH, Zhu T, Zhang YM, Zheng XX, Zhang TF, Lin YY, Wu ZY, Liu ZY, Lin Y, et al. Noninvasive artificial intelligence system for early predicting residual cancer burden during neoadjuvant chemotherapy in breast cancer. Annals of surgery; 2024.
- Huang Y, Zhu T, Zhang X, Li W, Zheng X, Cheng M, Ji F, Zhang L, Yang C, Wu Z et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. eClinicalMedicine. 2023;58.
- Li W, Partridge SC, Newitt DC, Steingrimsson J, Marques HS, Bolan PJ, Hirano M, Bearce BA, Kalpathy-Cramer J, Boss MA, et al. Breast multiparametric MRI for prediction of neoadjuvant chemotherapy response in breast cancer: the BMMR2 challenge. Radiol Imaging cancer. 2024;6(1):e230033.
- 12. Vaswani A. Attention is all you need. Adv Neural Inf Process Syst (2017).
- Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, Bolan PJ, Marques HS, Romanoff J, Cimino L, et al. Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: the ACRIN 6698 multicenter trial. Radiology. 2018;289(3):618–27.
- 14. Gradishar WJ, Moran MS, Abraham J, Abramson V, Aft R, Agnese D, Allison KH, Anderson B, Burstein HJ, Chew H, et al. NCCN Guidelines® insights:

breast cancer, version 4.2023. J Natl Compr Cancer Network: JNCCN. 2023;21(6):594–608.

- von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, Gerber B, Eiermann W, Hilfrich J, Huober J, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncology: Official J Am Soc Clin Oncol. 2012;30(15):1796–804.
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. Userguided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage. 2006;31(3):1116–28.
- Huang YH, Shi ZY, Zhu T, Zhou TH, Li Y, Li W, Qiu H, Wang SQ, He LF, Wu ZY et al. Longitudinal MRI-Driven Multi-Modality Approach for Predicting Pathological Complete Response and B Cell Infiltration in Breast Cancer. Advanced science (Weinheim, Baden-Wurttemberg, Germany). 2025;e2413702.
- Zhang Y, You C, Pei Y, Yang F, Li D, Jiang YZ, Shao Z. Integration of radiogenomic features for early prediction of pathological complete response in patients with triple-negative breast cancer and identification of potential therapeutic targets. J Translational Med. 2022;20(1):256.
- Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, Gweon HM. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. Radiology. 2020;294(1):31–41.
- van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H. Computational radiomics system to Decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–7.
- Braman N, Prasanna P, Whitney J, Singh S, Beig N, Etesami M, Bates DDB, Gallagher K, Bloch BN, Vulchi M, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2 (ERBB2)-Positive breast cancer. JAMA Netw Open. 2019;2(4):e192561.
- Ding J, Chen S, Serrano Sosa M, Cattell R, Lei L, Sun J, Prasanna P, Liu C, Huang C. Optimizing the peritumoral region size in radiomics analysis for Sentinel lymph node status prediction in breast cancer. Acad Radiol. 2022;29:S223–8.
- 23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
- @article{adaloglou2021transformer. Title = Transformers in computer vision, author = Adaloglou, Nikolas, journal = Https://theaisummer.com/, year = 2021, howpublished = {https://github.com/The-Al-Summer/self-attention-c v}.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
- Park YH, Lal S, Lee JE, Choi YL, Wen J, Ram S, Ding Y, Lee SH, Powell E, Lee SK, et al. Chemotherapy induces dynamic immune responses in breast cancers that impact treatment outcome. Nat Commun. 2020;11(1):6175.

- Du S, Gao S, Zhao R, Liu H, Wang Y, Qi X, Li S, Cao J, Zhang L. Contrast-free MRI quantitative parameters for early prediction of pathological response to neoadjuvant chemotherapy in breast cancer. Eur Radiol. 2022;32(8):5759–72.
- Fan M, Chen H, You C, Liu L, Gu Y, Peng W, Gao X, Li L. Radiomics of tumor heterogeneity in longitudinal dynamic Contrast-Enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. Front Mol Biosci. 2021;8:622219.
- Gao Y, Ventura-Diaz S, Wang X, He M, Xu Z, Weir A, Zhou HY, Zhang T, van Duijnhoven FH, Han L, et al. An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. Nat Commun. 2024;15(1):9613.
- Zeng Q, Xiong F, Liu L, Zhong L, Cai F, Zeng X. Radiomics based on DCE-MRI for predicting response to neoadjuvant therapy in breast cancer. Acad Radiol. 2023;30(Suppl 2):S38–49.
- Hoffmann E, Gerwing M, Krähling T, Hansen U, Kronenberg K, Masthoff M, Geyer C, Höltke C, Wachsmuth L, Schinner R, et al. Vascular response patterns to targeted therapies in murine breast cancer models with divergent degrees of malignancy. Breast cancer Research: BCR. 2023;25(1):56.
- Jardim-Perassi BV, Huang S, Dominguez-Viqueira W, Poleszczuk J, Budzevich MM, Abdalah MA, Pillai SR, Ruiz E, Bui MM, Zuccari DAPC, et al. Multiparametric MRI and coregistered histology identify tumor habitats in breast cancer mouse models. Cancer Res. 2019;79(15):3952–64.
- Nuciforo P, Pascual T, Cortés J, Llombart-Cussac A, Fasani R, Paré L, Oliveira M, Galvan P, Martínez N, Bermejo B, et al. A predictive model of pathologic response based on tumor cellularity and tumor-infiltrating lymphocytes (CeITIL) in HER2-positive breast cancer treated with chemo-free dual HER2 Blockade. Annals Oncology: Official J Eur Soc Med Oncol. 2018;29(1):170–7.
- Han X, Guo Y, Ye H, Chen Z, Hu Q, Wei X, Liu Z, Liang C. Development of a machine learning-based radiomics signature for estimating breast cancer TME phenotypes and predicting anti-PD-1/PD-L1 immunotherapy response. Breast cancer Research: BCR. 2024;26(1):18.
- Sahoo S, Lester SC. Pathology of breast carcinomas after neoadjuvant chemotherapy: an overview with recommendations on specimen processing and reporting. Arch Pathol Lab Med. 2009;133(4):633–42.
- 36. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. Radiology. 2021;298(3):505–16.
- 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.

# Publisher's note

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