Radiomics-Guided Multi-Modal Learning for Pathological Complete Response Prediction from Breast MRI Under Incomplete Modality Settings

Xinyuan Xiang , Wenyu Yin , Jiayue Li * , Shufang Li

Posted Date: 23 April 2026

doi: 10.20944/preprints202604.1604.v1

Keywords: breast cancer; pCR prediction; multi-modal fusion; radiomics; incomplete modalities

 doi:10.20944/preprints202604.1604.v1

Article

Radiomics-Guided Multi-Modal Learning for Pathological Complete Response Prediction from Breast MRI Under Incomplete Modality Settings

Xinyuan Xiang [1], Wenyu Yin [1], Jiayue Li [2,*] and Shufang Li [1]

[1] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China
[2] School of Artificial Intelligence and Computer Science, North China University of Technology, Beijing, China
* Correspondence: lijy2017@ncut.edu.cn

Abstract

Pathological complete response (pCR) after neoadjuvant therapy is an important indicator of treatment response and prognosis in breast cancer. Multi-modal breast MRI provides complementary information for pCR prediction, but existing methods often assume complete modality availability and do not fully exploit the complementary value of radiomics and deep features. To address these limitations, we propose a radiomics-guided multi-modal learning framework for pCR prediction from breast MRI under incomplete modality settings. The model employs a multi-branch 2.5D encoder to extract modality-specific features, a radiomics-guided gating module to enhance deep representations with handcrafted priors, and a masked fusion strategy to adaptively integrate available modalities while excluding missing ones. Experiments on the I-SPY1 Trial dataset show that the proposed method achieves promising performance and maintains robustness under incomplete modality settings. These results suggest that the proposed framework can effectively integrate multi-modal MRI and radiomics information for pCR prediction and shows potential under incomplete modality settings.

Keywords: breast cancer; pCR prediction; multi-modal fusion; radiomics; incomplete modalities

1. Introduction

Breast cancer remains one of the most common malignancies among women worldwide and continues to pose a major challenge to global health. Neoadjuvant therapy (NAT) has become an important treatment strategy for patients with locally advanced or biologically aggressive breast cancer, as it can reduce tumor burden before surgery and provide an opportunity for early assessment of treatment response [1]. Among the available response indicators, pathological complete response (pCR) is widely recognized as a clinically meaningful endpoint because it is associated with improved long-term outcomes in specific breast cancer subtypes [2]. Therefore, accurate preoperative prediction of pCR is of substantial clinical value for treatment planning, response monitoring, and individualized therapeutic decision-making.

MRI plays a central role in breast cancer diagnosis and treatment evaluation owing to its superior soft-tissue contrast and multi-parametric imaging capability. In particular, breast MRI has shown considerable value in assessing residual disease and pathological response after NAT [3]. Multi-modal breast MRI, including dynamic contrast-enhanced (DCE) MRI, T1-weighted imaging, and T2-weighted imaging, provides complementary information regarding tumor morphology, vascularity, internal heterogeneity, and surrounding tissue characteristics. Benefiting from these rich imaging representations, recent studies have increasingly explored quantitative imaging analysis methods for treatment response prediction from breast MRI and have reported encouraging results [4–8].

In parallel with the development of quantitative imaging analysis, radiomics has emerged as an important imaging biomarker paradigm. By extracting handcrafted descriptors of lesion shape, intensity distribution, and texture heterogeneity, radiomics can provide complementary and relatively

 doi:10.20944/preprints202604.1604.v1

interpretable information beyond visual assessment [9–11]. Several studies have shown that radiomics-based or radiomics-assisted models can achieve promising performance in breast cancer response prediction [6–8]. These advances suggest that handcrafted quantitative descriptors and data-driven representation learning may offer complementary strengths for building more effective predictive models.

Despite recent progress, two limitations remain. First, many multi-modal methods assume that all MRI modalities are available during both training and inference. This assumption is often unrealistic in clinical practice, where incomplete MRI examinations are common. As a result, these models may show limited robustness and reduced clinical applicability when some modalities are missing. Second, although radiomics features provide quantitative and clinically meaningful priors, many studies simply concatenate them with deep features. This strategy does not explicitly model how radiomics can guide deep feature learning and thus may not fully exploit the complementarity between handcrafted and deep representations.

To address these issues, we propose a radiomics-guided multi-modal learning framework for pathological complete response prediction from breast MRI under incomplete modality settings. The framework uses a multi-branch 2.5D encoder to extract modality-specific features from DCE, T1-weighted, and T2-weighted images. A radiomics-guided gating module is introduced to recalibrate deep features using handcrafted radiomic priors. In addition, a masked fusion strategy is designed to aggregate only the available modalities, thereby improving robustness to missing inputs. The main contributions of this study are summarized as follows:

- We propose a radiomics-guided multi-modal learning framework for pCR prediction from breast MRI, which jointly exploits deep imaging features and handcrafted radiomics information.
- We develop a radiomics-guided feature modulation mechanism that enables clinically relevant radiomics priors to enhance modality-specific deep representations.
- We design a masked fusion strategy for incomplete modality settings, which enables the model to aggregate available modality information in a flexible manner and enhances its adaptability when one or more modalities are missing.

Experiments on the I-SPY1 Trial dataset show that the proposed method achieves competitive predictive performance and yields encouraging results under different modality availability settings.

2. Methodology

2.1. Overview

The overall framework of RGM-Net is illustrated in Figure 1. The proposed model is designed for pathological complete response prediction from multi-modal breast MRI under incomplete modality settings. Let $x = \{X^{(m)}\}$, $m \in [1,2,3]$ denote the set of available MRI modalities, where $X^{(m)}$ is the input image of the m-th modality. In this study, the considered modalities include DCE MRI, T1-weighted imaging, and T2-weighted imaging. In addition, a radiomics feature vector R is extracted from the lesion region to provide complementary handcrafted information.

 doi:10.20944/preprints202604.1604.v1

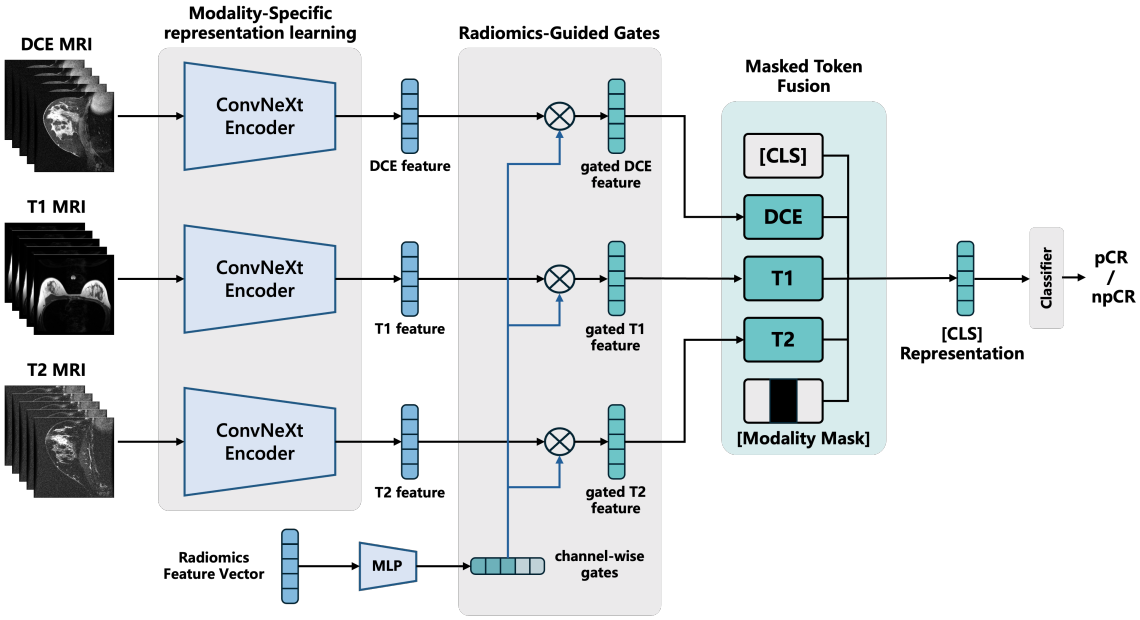

Figure 1. Overall framework of the proposed RGM-Net.

The framework consists of three main components: a multi-branch 2.5D feature encoder, a radiomics-guided gating (RGG) module, and a masked token fusion (MTF) module. First, each modality is processed by an independent branch to extract modality-specific deep features. Next, the RGG module transforms the radiomics features into channel-wise weights to recalibrate the deep representations. Finally, the MTF module performs modality-aware fusion using only the available modality tokens, thereby preventing missing modalities from interfering with feature aggregation. The fused representation is then fed into the prediction head for pCR classification.

2.2. Multi-Branch Feature Encoding

Multi-modal breast MRI contains heterogeneous yet complementary information across different imaging sequences. To preserve modality-specific characteristics, each modality is processed by a dedicated branch with the same network architecture but independent parameters. A 2.5D ConvNeXt backbone is adopted as the feature extractor for each branch.

For the m-th modality, five adjacent slices centered on the lesion are stacked along the channel dimension to form a 2.5D input. This design provides richer contextual information than single-slice 2D analysis while avoiding the high computational cost and data demand of full 3D convolution. Given the input $X^{(m)} \in \mathbb{R}^{C\times H\times W}$, the encoded deep feature map $F^{(m)}$ is obtained as

$$F^{(m)} = E^{(m)}\left(X^{(m)}\right), \tag{1}$$

where $E^{(m)}(\cdot)$ denotes the ConvNeXt encoder corresponding to the m-th modality. By using separate encoding branches, the model can learn modality-specific representations tailored to the imaging characteristics of DCE, T1-weighted, and T2-weighted MRI, thereby providing a stronger basis for subsequent cross-modal fusion.

2.3. Radiomics-Guided Gating

Although deep convolutional features can capture hierarchical semantic representations, they may not explicitly preserve clinically meaningful quantitative characteristics. Radiomics features, in contrast, provide handcrafted descriptors of lesion morphology, intensity distribution, and textural heterogeneity, which offer complementary and relatively interpretable prior information. To better integrate such prior knowledge into deep representation learning, we introduce a RGG module.

 doi:10.20944/preprints202604.1604.v1

Given the radiomics feature vector R, a multi-layer perceptron (MLP) is used to project it into a channel-wise gating vector:

$$G = \sigma(\mathrm{MLP}(R)), \tag{2}$$

where G denotes the learned gating weights and $\sigma(\cdot)$ denotes the sigmoid function. The output values are constrained to $[0, 1]$, enabling the model to adaptively control the importance of different feature channels. The gating vector is then applied to the encoded deep feature map through channel-wise scaling:

$$\widetilde{F}^{(m)} = G \odot F^{(m)}, \tag{3}$$

where $\odot$ denotes element-wise multiplication, and $\widetilde{F}^{(m)}$ is the radiomics-modulated feature map of the m-th modality.

In this manner, the radiomics-guided gating module serves as an explicit bridge between hand-crafted clinical priors and latent deep representations. Rather than simply concatenating radiomics and deep features, the proposed strategy uses radiomics information to recalibrate feature responses in a targeted channel-wise manner, thereby emphasizing clinically relevant patterns and suppressing less informative feature channels.

2.4. Masked Token Fusion

To address incomplete modality settings, a masked token fusion strategy is introduced. For each modality m, the radiomics-modulated feature map $\widetilde{F}^{(m)}$ is first globally pooled and then projected by a linear layer into a modality token $T^{(m)} \in \mathbb{R}^{128}$:

$$T^{(m)} = \mathcal{P}\left(\widetilde{F}^{(m)}\right), \tag{4}$$

where $\mathcal{P}(\cdot)$ denotes the projection function composed of global average pooling and a linear layer. A learnable [CLS] token is prepended to the token sequence for global representation aggregation. A modality mask $M^{(m)} \in \{0, 1\}$ is defined to indicate whether modality m is available. Based on the projected tokens and modality mask, feature fusion is performed through mask-aware attention:

$$\hat{T} = \mathrm{Attention}\left(T^{(1)}, T^{(2)}, T^{(3)}; M\right). \tag{5}$$

In the attention computation, the mask assigns a value of $-\infty$ to the logits associated with missing modalities ($M^{(m)} = 0$). After the softmax operation, these positions receive zero attention weights. Therefore, unavailable modalities do not contribute to feature aggregation. In this way, the [CLS] token only integrates information from available modality tokens.

The fused global representation is then passed through a Layer Normalization layer and a linear classifier to produce the final pCR prediction. This design enables adaptive fusion under incomplete inputs, avoids the need for data imputation, and improves robustness in real-world clinical settings.

2.5. Loss Function

The model is trained using cross-entropy loss and focal loss. Cross-entropy loss provides standard supervision, while focal loss emphasizes hard samples and reduces the influence of class imbalance. The losses are defined as

$$L_{\mathrm{ce}} = -[y \log(p) + (1 - y) \log(1 - p)], \tag{6}$$

$$L_{\mathrm{focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t), \tag{7}$$

where y is the ground-truth label, p is the predicted probability for the positive class, and $p_t = p$ for positive samples and $p_t = 1 - p$ for negative samples. α and γ denote the weighting and focusing parameters, respectively. The total loss is

$$\mathcal{L} = L_{\mathrm{ce}} + L_{\mathrm{focal}}. \tag{8}$$

 doi:10.20944/preprints202604.1604.v1

3. Experiments and Results

3.1. Data Acquisition and Preprocessing

The proposed model was evaluated on the I-SPY1 Trial dataset, which included 157 patients, comprising 43 pCR cases and 114 non-pCR cases. All data splits were performed at the patient level to avoid information leakage. The dataset was divided into training, validation, and test sets at a ratio of 7:2:1.

A 2.5D training strategy was adopted to capture inter-slice contextual information while maintaining computational efficiency. For each patient, the central lesion slice and its four adjacent slices were stacked to form a five-channel input, and all images were resized to 224×224 pixels. To simulate incomplete modality settings during training, the T1-weighted and T2-weighted modalities were each randomly masked with a probability of 0.5, while the DCE modality was always retained. In addition, 107 radiomic features were extracted using the PyRadiomics library. The lesion ROI was determined according to the tumor region in the DCE modality provided in the original dataset and was used for radiomics feature extraction only, without involving pathological outcome labels during preprocessing.

3.2. Implementation Details

The proposed framework was implemented in PyTorch. ConvNeXt was used as the backbone for each modality-specific branch. The model was optimized using the AdamW optimizer with an initial learning rate of 1×10^{-4} and a weight decay of 0.05. A cosine annealing scheduler was employed to adjust the learning rate during training. The model was trained for 400 epochs with a batch size of 32. To improve generalization, data augmentation methods, including horizontal flipping and random rotation, were applied during training. All experiments were conducted on an NVIDIA A6000 GPU.

3.3. Evaluation Metrics

The classification performance was assessed using accuracy, sensitivity, specificity, precision, and F1-score. These metrics were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{10}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{12}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}, \tag{13}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

3.4. Performance and Baseline Comparison

We compared the proposed method with several representative baseline models, including ResNet [12], ConvNeXt [13], EfficientNetV2 [14], and Vision Transformer (ViT) [15], under single-modality settings. In addition, the proposed multi-modal framework was evaluated using all available modalities. The quantitative results are summarized in Table 1.

Among the single-modality settings, DCE consistently showed the best performance across different backbones. Specifically, DCE achieved an accuracy of 81.25% under ConvNeXt, EfficientNetV2, and ViT, outperforming the corresponding T1-weighted and T2-weighted settings. This finding suggests that DCE-MRI provides the most discriminative information for pCR prediction among the three

 doi:10.20944/preprints202604.1604.v1

modalities. By contrast, T1-weighted and T2-weighted images yielded relatively lower performance, indicating that either modality alone may not be sufficient to comprehensively characterize treatment response.

When all modalities were jointly incorporated, the proposed method achieved the best overall performance, with an accuracy of 86.25%, a sensitivity of 77.50%, a specificity of 95.00%, a precision of 93.94%, and an F1-score of 84.93%. The confusion matrices shown in Figure 2 further demonstrate the classification characteristics of the proposed method. In particular, the high specificity and precision indicate that the model is effective in identifying non-pCR cases, while the sensitivity remains competitive for detecting pCR cases.

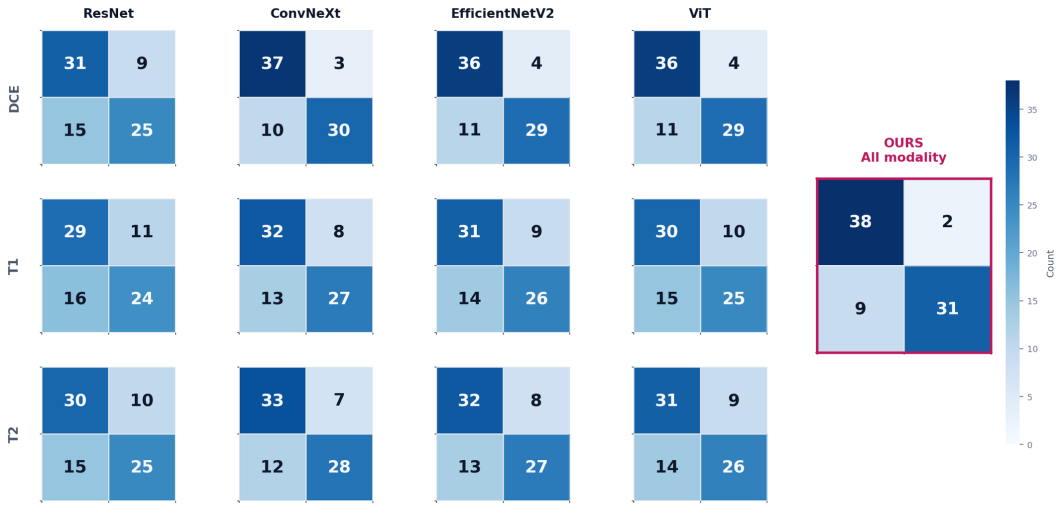

Figure 2. Confusion matrices of different methods.

Table 1. Comparison of different methods.

Method	Modality	Acc.	Sen.	Spe.	Pre.	F1
ResNet	DCE	70.00	62.50	77.50	73.53	67.57
	T1	66.25	60.00	72.50	68.57	64.00
	T2	68.75	62.50	75.00	71.43	66.67
ConvNeXt	DCE	81.25	72.50	90.00	87.88	79.45
	T1	73.75	67.50	80.00	77.14	72.00
	T2	76.25	70.00	82.50	80.00	74.67
EffNetV2	DCE	81.25	72.50	90.00	87.88	79.45
	T1	71.25	65.00	77.50	74.29	69.33
	T2	73.75	67.50	80.00	77.14	72.00
ViT	DCE	81.25	72.50	90.00	87.88	79.45
	T1	68.75	62.50	75.00	71.43	66.67
	T2	71.25	65.00	77.50	74.29	69.33
OURS	**All Modalities**	**86.25**	**77.50**	**95.00**	**93.94**	**84.93**

3.5. Ablation Study

To investigate the contribution of the proposed modules, we conducted an ablation study by comparing the baseline ConvNeXt model, the multi-modal model without radiomics guidance, and the full proposed framework. The results are summarized in Table 2.

Compared with the baseline single-modality ConvNeXt model, the multi-modal framework without radiomics guidance already improved predictive performance, indicating that multi-modal feature fusion is beneficial for pCR prediction. After further incorporating the radiomics-guided gating

 doi:10.20944/preprints202604.1604.v1

module, the full model achieved the best results across all evaluation metrics. Specifically, the full model improved the accuracy from 81.25% to 86.25% and the F1-score from 79.45% to 84.93%.

These results suggest that the proposed framework benefits from both multi-modal fusion and radiomics-guided feature modulation. The observed improvement supports the hypothesis that handcrafted radiomics descriptors provide useful complementary priors for refining deep feature learning.

Table 2. Ablation study of the proposed modules.

Method	Modality	Acc.	Sen.	Spe.	Pre.	F1
ConvNeXt (Base.)	DCE	81.25	72.50	90.00	87.88	79.45
	T1	73.75	67.50	80.00	77.14	72.00
	T2	76.25	70.00	82.50	80.00	74.67
w/o Radiomics Guidance	All Modalities	83.75	75.00	92.50	90.91	82.19
OURS	**All Modalities**	**86.25**	**77.50**	**95.00**	**93.94**	**84.93**

3.6. Modality Analysis

To further investigate the impact of different modality combinations, we evaluated the proposed framework under four settings: DCE alone, DCE+T1, DCE+T2, and all modalities. The results are presented in Table 3.

Among the single- and dual-modality settings, DCE alone already achieved strong performance, with an accuracy of 82.50% and an F1-score of 81.08%, highlighting its important role in pCR prediction. Incorporating an additional modality further improved the overall performance, indicating that T1-weighted and T2-weighted images provide complementary information beyond DCE. Specifically, DCE+T1 outperformed DCE+T2 in terms of accuracy (85.00% vs. 83.75%), sensitivity (75.00% vs. 72.50%), and F1-score (83.33% vs. 81.69%), while both combinations achieved the same specificity of 95.00%.

When all modalities were jointly used, the proposed framework achieved the best overall performance, with an accuracy of 86.25%, a sensitivity of 77.50%, and an F1-score of 84.93%. These results suggest that integrating multiple MRI modalities can improve predictive performance by leveraging complementary information from different imaging sequences.

Table 3. Performance under different modality settings.

Modality	Acc.	Sen.	Spe.	Pre.	F1
DCE	82.50	75.00	90.00	88.24	81.08
DCE+T1	85.00	75.00	95.00	93.75	83.33
DCE+T2	83.75	72.50	95.00	93.55	81.69
All Modalities	**86.25**	**77.50**	**95.00**	**93.94**	**84.93**

4. Discussion

This study proposed a radiomics-guided multi-modal learning framework for pathological complete response prediction from breast MRI under incomplete modality settings. The experimental results showed that the proposed method achieved better overall performance than the baseline models and remained effective under different modality combinations. These findings support our hypothesis that deep imaging features and handcrafted radiomics information provide complementary value for pCR prediction.

An important result of this study is that DCE-MRI consistently achieved the best performance among the single-modality settings. This observation is consistent with previous studies showing that DCE-MRI contains rich information on tumor vascularity, morphology, and treatment-related changes. These characteristics are closely related to treatment response. At the same time, the

 doi:10.20944/preprints202604.1604.v1

modality analysis showed that adding T1-weighted or T2-weighted imaging further improved the predictive performance compared with DCE alone. This suggests that different MRI sequences provide complementary information and that multi-modal fusion can improve tumor characterization.

The ablation study further demonstrated the effectiveness of the proposed radiomics-guided gating module. Compared with the multi-modal model without radiomics guidance, the full model achieved better performance across all evaluation metrics. This result indicates that radiomics features can provide useful prior information for refining deep feature learning. Instead of simply concatenating radiomics and deep features, the proposed method uses radiomics descriptors to recalibrate deep feature channels in a targeted manner. This design may help the model emphasize more relevant patterns and suppress less informative responses.

Another important aspect of the proposed framework is its ability to handle incomplete modality settings. In real clinical practice, complete multi-modal MRI data are not always available because of differences in acquisition protocols, motion artifacts, or missing imaging sequences. Many existing multi-modal methods assume complete inputs during both training and inference. Such an assumption may limit their clinical applicability. In contrast, the proposed masked token fusion strategy aggregates only the available modalities and prevents missing inputs from interfering with feature fusion. The encouraging results under different modality combinations suggest that this strategy improves the flexibility and robustness of the model.

Several limitations should also be noted. First, the study was conducted on a relatively small public dataset, which may limit the generalizability of the findings. Second, the simulated incomplete modality setting may not fully reflect the complexity of missing-data patterns in real clinical scenarios. Third, although radiomics guidance improved performance, the interpretability of the interaction between radiomics priors and deep representations still needs further investigation.

Future work can be extended in several directions. First, the proposed framework should be validated on larger multi-center datasets. Second, more realistic missing-modality learning strategies may be explored. Third, additional information, such as clinical variables and molecular subtype, may be incorporated to further improve prediction performance. In addition, improving model interpretability would be important for future clinical application.

Overall, the results suggest that radiomics-guided multi-modal learning is a promising strategy for breast MRI-based pCR prediction. By combining multi-modal imaging information, handcrafted radiomics priors, and incomplete-modality modeling, the proposed framework provides an effective solution for treatment response prediction in breast cancer.

5. Conclusions

This study proposed a radiomics-guided multi-modal learning framework for pathological complete response prediction from breast MRI under incomplete modality settings. By integrating modality-specific deep features and radiomics information, the proposed method was able to exploit complementary information from multiple MRI sequences. Experimental results on the I-SPY1 Trial dataset demonstrated encouraging predictive performance. These findings suggest that the proposed framework is a promising approach for pCR prediction and may provide improved flexibility when dealing with incomplete modality inputs.

Author Contributions: Conceptualization, X.X. and J.L.; methodology, X.X.; software, X.X.; validation, X.X. and W.Y.; formal analysis, X.X.; investigation, X.X.; resources, J.L. and S.L.; data curation, X.X. and W.Y.; writing—original draft preparation, X.X.; writing—review and editing, J.L. and S.L.; visualization, X.X.; supervision, J.L. and S.L.; project administration, J.L.; funding acquisition, J.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62506008, the R&D Program of Beijing Municipal Education Commission under Grant 110052972508-10, the Youth Research Special Project of North China University of Technology (NCUT) under Grant 2025NCUTYRSP014, and the Scientific Research Start-up Fund under Project No. 11005136025XN076-046.

 doi:10.20944/preprints202604.1604.v1

Institutional Review Board Statement: Not applicable. The study used publicly available de-identified data and did not involve direct human or animal participation.

Informed Consent Statement: Not applicable. The study used publicly available de-identified data.

Data Availability Statement: The datasets analyzed during the current study are publicly available in the I-SPY1 Trial dataset repository on Kaggle: https://www.kaggle.com/datasets/saarthakkapse/ispy1-trail-dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DCE	Dynamic Contrast-Enhanced
MRI	Magnetic Resonance Imaging
MLP	Multi-Layer Perceptron
NAT	Neoadjuvant Therapy
non-pCR	Non-Pathological Complete Response
pCR	Pathological Complete Response
ROI	Region of Interest

References

1. Montemurro, F.; Nuzzolese, I.; Ponzone, R. Neoadjuvant or adjuvant chemotherapy in early breast cancer? *Expert opinion on pharmacotherapy* **2020**, *21*, 1071–1082.
2. Von Minckwitz, G.; Untch, M.; Blohmer, J.U.; Costa, S.D.; Eidtmann, H.; Fasching, P.A.; 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. *Journal of clinical oncology* **2012**, *30*, 1796–1804.
3. Lobbes, M.; Prevos, R.; Smidt, M.; Tjan-Heijnen, V.; Van Goethem, M.; Schipper, R.; Beets-Tan, R.; Wildberger, J. The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: a systematic review. *Insights into imaging* **2013**, *4*, 163–175.
4. Huang, S.y.; Franc, B.L.; Harnish, R.J.; Liu, G.; Mitra, D.; Copeland, T.P.; Arasu, V.A.; Kornak, J.; Jones, E.F.; Behr, S.C.; et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ breast cancer* **2018**, *4*, 24.
5. Chitalia, R.D.; Rowland, J.; McDonald, E.S.; Pantalone, L.; Cohen, E.A.; Gastounioti, A.; Feldman, M.; Schnall, M.; Conant, E.; Kontos, D. Imaging phenotypes of breast cancer heterogeneity in preoperative breast dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) scans predict 10-year recurrence. *Clinical Cancer Research* **2020**, *26*, 862–869.
6. Xia, B.; Wang, H.; Wang, Z.; Qian, Z.; Xiao, Q.; Liu, Y.; Shao, Z.; Zhou, S.; Chai, W.; You, C.; et al. A combined nomogram model to predict disease-free survival in triple-negative breast cancer patients with neoadjuvant chemotherapy. *Frontiers in Genetics* **2021**, *12*, 783513.
7. Ma, M.; Gan, L.; Liu, Y.; Jiang, Y.; Xin, L.; Liu, Y.; Qin, N.; Cheng, Y.; Liu, Q.; Xu, L.; et al. Radiomics features based on automatic segmented MRI images: prognostic biomarkers for triple-negative breast cancer treated with neoadjuvant chemotherapy. *European Journal of Radiology* **2022**, *146*, 110095.
8. Zhang, L.; Jiang, X.; Xie, X.; Wu, Y.; Zheng, S.; Tian, W.; Xie, X.; Li, L. The impact of preoperative radiomics signature on the survival of breast cancer patients with residual tumors after NAC. *Frontiers in Oncology* **2021**, *10*, 523327.
9. Gillies, R.J.; Anderson, A.; Gatenby, R.; Morse, D. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clinical radiology* **2010**, *65*, 517–521.
10. Mayerhoefer, M.E.; Materka, A.; Langs, G.; Häggström, I.; Szczypiński, P.; Gibbs, P.; Cook, G. Introduction to radiomics. *Journal of Nuclear Medicine* **2020**, *61*, 488–495.
11. Guiot, J.; Vaidyanathan, A.; Deprez, L.; Zerka, F.; Danthine, D.; Frix, A.N.; Lambin, P.; Bottari, F.; Tsoutzidis, N.; Miraglio, B.; et al. A review in radiomics: making personalized medicine a reality via routine imaging. *Medicinal research reviews* **2022**, *42*, 426–440.
12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

 doi:10.20944/preprints202604.1604.v1

13. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11976–11986.
14. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International conference on machine learning. PMLR, 2021, pp. 10096–10106.
15. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* **2020**.