Longitudinal MRI-Clinical Multimodal Fusion for pCR Prediction in Breast Cancer

Dingrui Ma[1,2], Jiawei Cao[1,2], Hao Cheng[1,2(✉)], Dan Zhou[3], Jianping Liu[4], Xiaofeng Zhang[1,2], Kaijie Wu[1,2], Chaochen Gu[1,2], and Xinping Guan[1,2]

[1] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
jiaodachenghao@sjtu.edu.cn
[2] Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China
[3] Department of Breast Surgery, Southern University of Science and Technology Affiliated Foshan Hospital (The First People's Hospital of Foshan), Foshan, Guangdong, China
[4] Department of Radiology, Southern University of Science and Technology Affiliated Foshan Hospital (The First People's Hospital of Foshan), Foshan, Guangdong, China

Abstract. Pathologic complete response (pCR) prediction for breast cancer patients undergoing neoadjuvant chemotherapy (NAC) is crucial for optimizing treatment strategies. Nowadays, an increasing number of studies focus on predicting NAC response using preoperative imaging, and with the advancement of deep learning, different modalities of imaging and other clinical data can be effectively integrated to provide more comprehensive information. However, existing deep learning methods primarily focus on multimodal fusion or longitudinal modeling but often suffer from inadequate feature focus and overlook specific treatment effects. To address these limitations, we propose a novel multimodal-learning framework LMF(**L**ongitudinal MRI-Clinical **M**ultimodal **F**usion) that enhances feature extraction and explicitly models treatment-induced imaging changes. Our method consists of two key components: (1) Molecular-Aware Deformable Attention (MADA), which integrates molecular subtype information with MRI features and refines spatial representations via deformable cross-attention mechanism; and (2) Treatment-Aware Longitudinal Modeling (TALM), which incorporates treatment embeddings to capture NAC-driven feature variations. The model is trained and evaluated on the ISPY-2 dataset, using pre- and post-NAC DCE-MRI alongside clinical data. Experimental results demonstrate that our approach outperforms existing methods, confirming that MADA effectively enhances feature extraction while TALM strengthens longitudinal modeling. These findings highlight the potential of integrating multimodal feature refinement with treatment-aware temporal modeling for improved pCR prediction. Our code is available at https://github.com/martin-bro/LMF.

Keywords: Multimodal Learning · Longitudinal Imaging Analysis · Neoadjuvant Chemotherapy · Pathologic Complete Response Prediction

1 Introduction

Neoadjuvant chemotherapy (NAC) improves surgical feasibility by shrinking tumor size preoperatively and provides a window to assess breast cancer biology [18]. Pathologic complete response (pCR) is a crucial NAC efficacy marker, strongly correlated with survival, particularly in HER2-positive and triple-negative subtypes [2]. However, current pCR evaluation relies on postoperative pathology, delaying treatment adjustments and limiting personalized strategies [20]. Early pCR prediction via preoperative imaging, such as DCE-MRI [17], offers the potential to optimize treatment plans and reduce overtreatment. Traditional machine learning methods depend on manual tumor delineation and handcrafted feature extraction, limiting their flexibility [7,8]. In contrast, deep learning enables high-dimensional feature extraction, surpassing handcrafted methods in multimodal fusion and longitudinal modeling [10].

Despite their advantages, deep learning methods face several challenges. Two-stage approaches (i.e., segmentation followed by handcrafted feature extraction) require highly accurate tumor delineation [15], while non-segmentation-based methods typically crop small regions around lesions during preprocessing [22,19]. However, the variable lesion sizes complicate optimal patch dimension selection. Direct processing of MRI scans covering most of the complete unilateral breast region leads to feature unfocused issues, necessitating more effective feature aggregation mechanisms. Emerging evidence suggests MRI characteristics correlate with breast cancer molecular subtypes [23], motivating **multimodal fusion with clinical data**. Although Du et al. [3] and Petersen et al. [16] proposed contrastive learning for image-tabular alignment, they focused on the entire image, lacking a method for integrating key features of MRI. Xiong et al. [21] introduced cross-attention for feature fusion, but their ablation studies revealed that the primary accuracy gains came from the feature dimension reduction module (CMLP) rather than the attention mechanism itself, suggesting cross-attention alone may be insufficient for efficient MRI-clinical data interaction.

Longitudinal imaging analysis leverages tumor dynamics during neoadjuvant chemotherapy (NAC) to enhance treatment response prediction. Recent advances include 3D RP-Net for MRI contrastive learning [1] and Siamese Multi-Task Networks (SMTN) for HER2-positive pCR prediction [13]. While Gao et al. [4] proposed timepoint encodings, their late fusion strategy inadequately captures complex feature transitions across treatment phases. Methods employing contrastive loss assumptions (e.g., Zhang et al. [22], Sun et al. [19]) posit greater imaging changes in pCR patients, but this hypothesis oversimplifies cases with subtle imaging distinctions. Furthermore, most existing approaches neglect explicit modeling of treatment-specific effects, potentially limiting model sensitivity to treatment variations. Recent explorations in treatment-effect modeling include Liu et al.'s treatment-aware diffusion for brain MRI [12] and Gao et al.'s NAC mammography synthesis [5]. However, the high anatomical variability in breast MRI constrains generative models' capacity to predict NAC-induced morphological changes, highlighting the need for alternative modeling paradigms.

To overcome these limitations, we propose a novel multimodal-learning framework LMF(**L**ongitudinal MRI-Clinical **M**ultimodal **F**usion) for pCR prediction with low annotation burden, which decouples the key steps of the breast cancer pCR prediction task (information extraction and longitudinal modeling). Our contributions are as follows.

1. We introduce Molecular-Aware Deformable Attention (MADA), a novel cross-attention mechanism, to integrate molecular subtype information with MRI features, enhancing feature extraction.
2. We propose Treatment-Aware Longitudinal Modeling (TALM), an explicit longitudinal modeling strategy, to incorporate treatment embeddings to capture changes induced by treatment, improving the prediction of pCR status from longitudinal imaging data.

2 Methods

2.1 Overall Structure

This study aims to predict the pathological complete response (pCR) of breast cancer patients undergoing neoadjuvant chemotherapy (NAC). The input data includes dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and clinical information. The model employs ResNet18-3D as the backbone for feature extraction, which is pre-trained using a segmentation network trained on a randomly split dataset (4:1) using pre-NAC MRI, and the annotations are from a benchmark for breast tumor segmentation [6] achieving a Dice coefficient of 0.7893. Additionally, a phaseConv module is introduced to model signal enhancement patterns between pre-contrast and Phase 3 images, enhancing the dynamic representation of DCE-MRI. Patients for segmentation model validation are not included in pCR prediction validation set.

Considering the weakly supervised nature of the dataset [9], where t0 represents the pre-NAC time point with only partial lesion annotations and t2 represents the post-NAC time point without precise lesion annotations, pseudo masks generated from the pre-trained segmentation network are used during training. Dice loss is applied to constrain the corresponding module, improving the learning of lesion-related features. The overall framework of the method is illustrated in Fig. 1, which consists of two key components: Molecular-Aware Deformable Attention (MADA) for multimodal feature refinement and Treatment-Aware Longitudinal Modeling (TALM) for treatment-sensitive longitudinal analysis.

2.2 Molecular-Aware Deformable Attention (MADA)

To address the issue of unfocused attention and redundant information when extracting large-field MRI features using 3D CNN, we propose Molecular-Aware Deformable Attention (MADA), which enhances MRI feature extraction through image-tabular multimodal fusion.

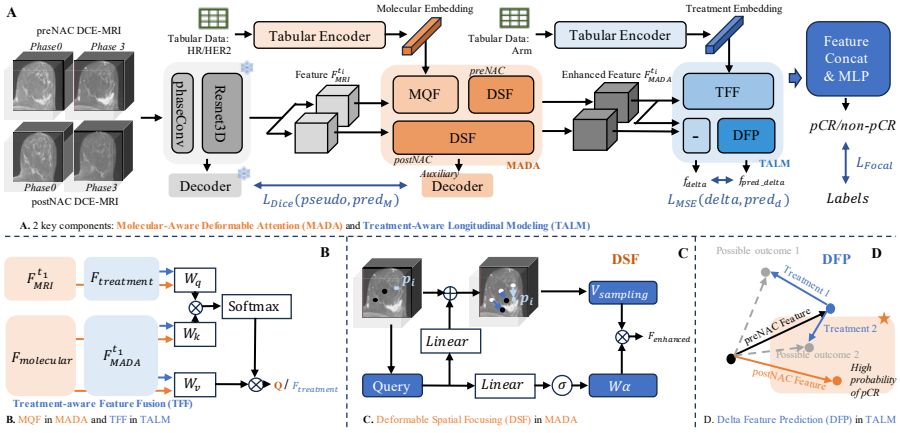

Fig. 1. Overall structure of our proposed model **LMF**. **A.** Our Model with 2 key components: Molecular-Aware Deformable Attention (MADA) and Treatment-Aware Longitudinal Modeling (TALM). **B.** MQF in MADA and TFF in TALM facilitate feature interaction between MRI, molecular, and treatment embeddings. **C.** DSF in MADA dynamically samples key regions in MRI to extract enriched feature representations. **D.** DFP in TALM models NAC-induced changes in MRI features.

Molecular-Aware Query Fusion We compute the interaction between molecular information and MRI features via Cross Attention to enhance the biological relevance of imaging features:

$$Q_{\text{multimodal}} = Q + \text{Softmax}\left(\frac{QK^T}{\sqrt{C}}\right)V, \tag{1}$$

where Query $Q \in \mathbb{R}^{B\times L_q\times C}$ is computed from MRI feature maps $F_{\text{MRI}} \in \mathbb{R}^{B\times C\times D\times H\times W}$ via flattening and a linear projection. Key $K \in \mathbb{R}^{B\times 1\times C}$ and Value $V \in \mathbb{R}^{B\times 1\times C}$ originate from molecular features $F_{\text{molecular}} \in \mathbb{R}^{B\times C}$, which are encoded via a Tabular Encoder (MLP-like) from raw clinical tabular data.

Deformable Spatial Focusing We obtain the molecular-integrated Query $Q_{multimodal}$ as the input to Deformable Attention [24], while Key and Value are mapped from MRI feature maps F_{MRI}. We first randomly initialize sampling points $p_i \in \mathbb{R}^{B_q\times L_q\times 1\times 3}$ and input them into the Deformable Attention Layer to obtain enhanced imaging features:

$$F_{\text{enhanced}} = \sum_{i=1}^{P} \text{Softmax}\left(\frac{QK^T}{\sqrt{C}}\right)_i \cdot F_{\text{MRI}}(p_i + \Delta p_i) = \mathcal{D}(Q, K, V, p), \tag{2}$$

where the offset $\Delta p_i = W_{\text{offsets}}Q$ is computed from a learnable offset prediction network. In implementation, we set the number of attention heads $n_{\text{heads}} = 8$ and the number of sampling points per head $n_{\text{points}} = 8$ to enhance model capacity.

We apply 3 iterative layers of Deformable Attention, where the output of each layer serves as the input for the next, enabling the progressive optimization of sampling locations. At the l-th layer, we directly use the enhanced imaging features from the previous layer, $F^{(l-1)}_{\text{enhanced}}$, as the new Query and compute the current layer's features:

$$F^{(l)}_{\text{enhanced}} = \mathcal{D}(F^{(l-1)}_{\text{enhanced}}, K, V, p). \tag{3}$$

Additionally, we update the feature representation using residual connections and refine the sampling point locations:

$$p_i^{(l)} = \text{sigmoid}(p_i^{(l-1)} + W^{(l)}_{\text{ref}} F^{(l)}_{\text{enhanced}}). \tag{4}$$

This strategy ensures that sampling points are iteratively optimized across different layers, allowing the final feature representation to focus more effectively on tumor regions. For the $t0$ and $t2$ time points, we apply Deformable Attention separately, where the Query at $t0$ adopts the Molecular-Aware Query, while at $t2$, the Query remains derived from the original MRI feature map. Then we get the enhanced features:

$$F^{t_i}_{\text{MADA}} = \mathcal{D}(Q_{t_i}, K_{t_i}, V_{t_i}, p_{t_i}), i = 0, 2. \tag{5}$$

Through Deformable Spatial Focusing, the network dynamically adjusts its receptive field during feature extraction, reducing irrelevant information and ensuring that the final MRI feature representation focuses more on tumor regions. Combined with Molecular-Aware Query Fusion, this deformable cross attention mechanism enhances the model's capability to capture the biological characteristics of tumors.

2.3 Treatment-Aware Longitudinal Modeling (TALM)

We propose Treatment-Aware Longitudinal Modeling (TALM), which explicitly models the effect of treatment information on imaging feature evolution to improve pCR prediction with the following two main components.

Treatment-Aware Feature Fusion Studies suggest that treatment regimens influence the evolution of tumor imaging characteristics [4]. To incorporate this information, we employ Cross Attention to compute the interaction between chemotherapy information and MRI features, ensuring that imaging features are adjusted based on treatment factors:

$$F_{\text{treatment}} = F_{\text{treatment}} + \text{Softmax}\left(\frac{QK^T}{\sqrt{C}}\right)V, \tag{6}$$

where the Query $Q \in \mathbb{R}^{B\times 1\times C}$ originates from molecular features $F_{\text{treatment}} \in \mathbb{R}^{B\times C}$, which is also encoded via a Tabular Encoder from raw clinical tabular data, while the Key and Value matrices are obtained from MRI features $F^{t0}_{\text{MADA}} \in \mathbb{R}^{B\times L_q\times C}$ via linear transformations. After Cross Attention computation, $F_{\text{treatment}}$ represents the treatment embedding refined by MRI features and is used in the subsequent modeling of imaging changes.

Delta Feature Prediction After Treatment-Aware Feature Fusion, we further employ supervised learning to model NAC-induced changes in imaging features. The true imaging change, or delta feature, is computed as:

$$\Delta_{\text{true}} = f^{t2}_{\text{MADA}} - f^{t0}_{\text{MADA}}, \tag{7}$$

where f^{t0}_{MADA} and f^{t2}_{MADA} are derived from F^{t0}_{MADA} and F^{t2}_{MADA} via a basic pooling module, respectively.

To predict the expected imaging change, we train an MLP to map treatment embeddings to delta features:

$$\Delta_{\text{pred}} = W_\delta F_{treatment}, \tag{8}$$

where W_δ is a learnable transformation matrix that projects the treatment embedding, refined by imaging features, into the delta feature space. To ensure the model learns how to predict NAC-induced changes from treatment embeddings, we apply an MSE loss to supervise Δ_{pred}:

$$L_{\text{delta}} = \text{MSE}(\Delta_{\text{true}}, \Delta_{\text{pred}}). \tag{9}$$

This loss function ensures that the model not only captures intrinsic imaging changes but also incorporates treatment information to model change dynamics.

Through TALM, the model jointly captures treatment information and longitudinal imaging changes.

3 Experiments

3.1 Implementation Details

The MRI input has a size of $64 \times 128 \times 128$ ($D \times H \times W$), which can cover most of the complete unilateral breast region. The dataset is sourced from ISPY-2 [14], and based on imaging completeness (availability of both pre-NAC and post-NAC time points, with DCE-MRI including Phase 0 and Phase 3), 707 patients were selected from an initial cohort of 985 patients. To ensure consistent MRI resolution, resampling was performed to achieve a pixel spacing of $1mm \times 1mm \times 2mm$ ($x \times y \times z$). Clinical data consists of HR/HER2 status (one-hot encoded with a dimension of 2) and chemotherapy regimen (Agent, one-hot encoded with a dimension of 12).

All experiments were conducted on two NVIDIA V100 GPUs. The model was trained for 100 epochs, with the first 5 epochs as a warm-up stage. The batch size was set to 8. The initial learning rate was 0.001 and was gradually reduced to 0.0001 using a cosine annealing schedule. The classification loss function was Focal Loss [11] to address class imbalance issues . The overall loss function is defined as:

$$L = L_{\text{focal}} + \lambda L_{\text{auxiliary}} = L_{\text{focal}} + \lambda_1 L_{\text{dice}} + \lambda_2 L_{\text{delta}}, \tag{10}$$

where λ balances all loss components during training. The model performance was evaluated using Area Under the Curve (AUC), Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE).

Table 1. Performance comparison of different model variants.

Model	Backbone	MADA		TALM	Evaluation			
		MQF	DSF		*AUC*↑	*ACC*↑	*SEN*	*SPE*
Baseline	✓	×	×	×	0.6064	0.7092	0.1277	**1.0**
Model1-1	✓	✓	×	×	0.7134	0.7376	0.5106	0.8511
Model1-2	✓	×	✓	×	0.6548	0.6596	0.4468	0.7660
Model1-3	✓	✓	✓	×	0.7372	0.7021	0.4894	0.8095
Model2-1	✓	×	×	✓	0.5715	0.6099	0.2979	0.7660
LMF	✓	✓	✓	✓	**0.7574**	**0.7518**	**0.6596**	0.7979

3.2 Results

Ablation Study We evaluate the impact of different model components on pCR prediction performance. Table 1 presents the performance comparison of different model variants.

Comparing the Baseline model (only Backbone and simple molecular embedding) with Model1-1, Model1-2, and Model1-3 demonstrates that incorporating MADA (including Molecular Query Fusion (MQF) and Deformable Spatial Focus (DSF)) significantly improves performance. Model1-1 and Model1-2 assess the independent contributions of MQF and DSF, showing that MQF alone (Model1-1) provides a notable boost, and Model1-3 indicates that combining MQF and DSF (Model1-3) further enhances predictive performance.

Next, the effect of the TALM module is examined by comparing Model2-1 and LMF. Model2-1, which incorporates TALM without MADA, shows a decline in AUC compared to the Baseline, despite some improvement in SEN. This suggests that without prior feature enhancement from MADA, the model struggles to extract meaningful information for effective delta feature modeling. In contrast, LMF, which integrates both MADA and TALM, achieves the best overall performance, with AUC reaching 0.7574 and substantial improvements in SEN (0.6596) and SPE (0.7979), demonstrating that their combination optimally enhances the model's predictive capability.

These results confirm that MADA enhances feature extraction, TALM improves longitudinal modeling, and their combination achieves optimal performance in pCR prediction.

Comparison with other Methods To validate the effectiveness of our approach, we compare it with existing methods in **MRI-Clinical Fusion (MCF)** and **Longitudinal Modeling (LM)**, Table 2 presents the results. For multimodal fusion, we evaluate naive concatenation (iMRrhpc) [4], Cross Attention (3DCT-ICH) [21], and our LMF incorporating MADA. The results indicate that MRI and tabular fusion significantly improves performance, confirming the importance of leveraging molecular and treatment information. Furthermore, deep fusion methods, especially our deformable cross-attention mechanism, outper-

Table 2. Comparison with other methods.

Method	Model Components		Evaluation			
	MCF	**LM**	*AUC*↑	*ACC*↑	*SEN*	*SPE*
3DCT-ICH [21]	✓	×	0.7134	0.7376	0.5106	**0.8511**
M2Fusion [22]	×	✓	0.7200	0.6809	0.4894	0.7766
iMRrhpc [4]	✓	✓	0.7205	0.7092	0.5106	0.8085
LMF (Ours)	✓	✓	**0.7574**	**0.7518**	**0.6596**	0.7979

form simple concatenation, demonstrating the benefits of structured multimodal feature integration.

For longitudinal modeling, we compare contrastive loss (M2Fusion) [22], iMRrhpc with timepoint embeddings [4], and our LMF with TALM. The experiments show that naive concatenation of timepoint embeddings struggles to model complex chemotherapy responses, while contrastive loss, although improving upon iMRrhpc, has limitations: it assumes that pCR cases exhibit large imaging changes while non-pCR cases remain stable, which does not always hold due to response heterogeneity. Additionally, contrastive loss focuses on feature alignment rather than explicitly modeling the causal effect of treatment, limiting its ability to capture NAC-induced imaging variations. Our full model (LMF) achieves the best performance, confirming that explicit treatment-aware modeling effectively captures longitudinal tumor dynamics.

4 Conclusions

In this work, we propose a novel LMF(**L**ongitudinal MRI-Clinical **M**ultimodal **F**usion) model for prediction of pathologic complete response (pCR) in breast cancer patients undergoing neoadjuvant chemotherapy (NAC). By decoupling the key steps of information extraction and temporal modeling, our approach improves prediction accuracy with low annotation burden. We introduced Molecular-Aware Deformable Attention (MADA) for enhanced feature extraction and Treatment-Aware Longitudinal Modeling (TALM) for modeling chemotherapy-induced changes. Trained and validated on the ISPY-2 dataset, our results demonstrate that the proposed modules effectively improve pCR prediction, outperforming traditional methods. This work highlights the potential of multimodal deep learning in improving treatment response assessment and pCR prediction in clinical practice.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62227811, 62473255, 62273235). Additional support was provided by the Deep Blue Program Fund Project of the Second Institute of Oceanography, Ministry of Natural Resources. The authors are also affiliated with the Key Laboratory for System Control and Information Processing, Ministry of Education of China and Shanghai Key Laboratory for Perception and Control in Industrial Network Systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, Y., Liu, Y., Wang, C., Elliott, M., Kwok, C.F., Peña-Solorzano, C., Tian, Y., Liu, F., Frazer, H., McCarthy, D.J., et al.: Braixdet: Learning to detect malignant breast lesion with incomplete annotations. Medical image analysis **96**, 103192 (2024)
2. Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J.P., 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. The Lancet **384**(9938), 164–172 (2014)
3. Du, S., Zheng, S., Wang, Y., Bai, W., O'Regan, D.P., Qin, C.: Tip: Tabular-image pre-training for multimodal classification with incomplete data. In: European Conference on Computer Vision. pp. 478–496. Springer (2024)
4. Gao, Y., Ventura-Diaz, S., Wang, X., He, M., Xu, Z., Weir, A., Zhou, H.Y., Zhang, T., van Duijnhoven, F.H., Han, L., et al.: An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. Nature Communications **15**(1), 9613 (2024)
5. Gao, Y., Zhou, H.Y., Wang, X., Zhang, T., Han, L., Lu, C., Liang, X., Teuwen, J., Beets-Tan, R., Tan, T., et al.: Improving neoadjuvant therapy response prediction by integrating longitudinal mammogram generation with cross-modal radiological reports: A vision-language alignment-guided model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 133–143. Springer (2024)
6. Garrucho, L., Reidel, C.A., Kushibar, K., Joshi, S., Osuala, R., Tsirikoglou, A., Bobowicz, M., del Riego, J., Catanese, A., Gwoździewicz, K., et al.: Mama-mia: A large-scale multi-center breast cancer dce-mri benchmark dataset with expert segmentations. arXiv preprint arXiv:2406.13844 (2024)
7. Gilad, M., Freiman, M.: Pd-dwi: Predicting response to neoadjuvant chemotherapy in invasive breast cancer with physiologically-decomposed diffusion-weighted mri machine-learning model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 36–45. Springer (2022)
8. 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 **58** (2023)
9. Jin, C., Yu, H., Ke, J., Ding, P., Yi, Y., Jiang, X., Duan, X., Tang, J., Chang, D.T., Wu, X., et al.: Predicting treatment response from longitudinal images using multi-task deep learning. Nature communications **12**(1), 1851 (2021)
10. Li, Z., Gao, J., Zhou, H., Li, X., Zheng, T., Lin, F., Wang, X., Chu, T., Wang, Q., Wang, S., et al.: Multiregional dynamic contrast-enhanced mri-based integrated system for predicting pathological complete response of axillary lymph node to neoadjuvant chemotherapy in breast cancer: Multicentre study. EBioMedicine **107** (2024)
11. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)

12. Liu, Q., Fuster-Garcia, E., Hovden, I.T., MacIntosh, B.J., Grødem, E.O., Brandal, P., Lopez-Mateu, C., Sederevičius, D., Skogen, K., Schellhorn, T., et al.: Treatment-aware diffusion probabilistic model for longitudinal mri generation and diffuse glioma growth prediction. IEEE Transactions on Medical Imaging (2025)
13. Liu, Y., Wang, Y., Wang, Y., Xie, Y., Cui, Y., Feng, S., Yao, M., Qiu, B., Shen, W., Chen, D., et al.: Early prediction of treatment response to neoadjuvant chemotherapy based on longitudinal ultrasound images of her2-positive breast cancer patients by siamese multi-task network: a multicentre, retrospective cohort study. EClinicalMedicine **52** (2022)
14. Newitt, D.C., Partridge, S., Zhang, Z., Gibbs, J., Chenevert, T., Rosen, M., Bolan, P., Marques, H., Romanoff, J., Cimino, L., et al.: Acrin 6698/i-spy2 breast dwi [data set]. The Cancer Imaging Archive **1** (2021)
15. Nitzan, S., Gilad, M., Freiman, M.: Automated prediction of breast cancer response to neoadjuvant chemotherapy from dwi data. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
16. Petersen, J.K., Licht, V., Nielsen, M., Munk, A.: Revisiting clip: Efficient alignment of 3d mri and tabular data using domain-specific foundation models. arXiv preprint arXiv:2501.14051 (2025)
17. Portnow, L.H., Kochkodan-Self, J.M., Maduram, A., Barrios, M., Onken, A.M., Hong, X., Mittendorf, E.A., Giess, C.S., Chikarmane, S.A.: Multimodality imaging review of her2-positive breast cancer and response to neoadjuvant chemotherapy. Radiographics **43**(2), e220103 (2023)
18. Spring, L.M., Bar, Y., Isakoff, S.J.: The evolving role of neoadjuvant therapy for operable breast cancer. Journal of the National Comprehensive Cancer Network **20**(6), 723–734 (2022)
19. Sun, Y., Li, K., Chen, D., Hu, Y., Zhang, S.: Lomia-t: A transformer-based longitudinal medical image analysis framework for predicting treatment response of esophageal cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 426–436. Springer (2024)
20. van der Voort, A., Louis, F.M., van Ramshorst, M.S., Kessels, R., Mandjes, I.A., Kemper, I., Agterof, M.J., van der Steeg, W.A., Heijns, J.B., van Bekkum, M.L., et al.: Mri-guided optimisation of neoadjuvant chemotherapy duration in stage ii–iii her2-positive breast cancer (train-3): a multicentre, single-arm, phase 2 study. The Lancet Oncology **25**(5), 603–613 (2024)
21. Xiong, Z., Zhao, K., Ji, L., Shu, X., Long, D., Chen, S., Yang, F.: Multi-modality 3d cnn transformer for assisting clinical decision in intracerebral hemorrhage. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 522–531. Springer (2024)
22. Zhang, S., Du, S., Sun, C., Li, B., Shao, L., Zhang, L., Wang, K., Liu, Z., Tian, J.: M2fusion: Multi-time multimodal fusion for prediction of pathological complete response in breast cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 458–468. Springer (2024)
23. Zhang, Y., Chen, J.H., Lin, Y., Chan, S., Zhou, J., Chow, D., Chang, P., Kwong, T., Yeh, D.C., Wang, X., et al.: Prediction of breast cancer molecular subtypes on dce-mri using convolutional neural network with transfer learning between two centers. European radiology **31**, 2559–2567 (2021)
24. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)