M2Fusion: Multi-time Multimodal Fusion for Prediction of Pathological Complete Response in Breast Cancer

Song Zhang[1,2], Siyao Du[3], Caixia Sun[4], Bao Li[5], Lizhi Shao[1,2], Lina Zhang[6], Kun Wang[7], Zhenyu Liu[1,2], and Jie Tian[1,4]

[1] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China zhenyu.liu@ia.ac.cn, jie.tian@ia.ac.cn
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3] The First Hospital of China Medical University, Shenyang, China
[4] School of Engineering Medicine, Beihang University, Beijing, China
[5] Center for Biomedical Imaging, University of Science and Technology of China, Hefei, China
[6] The Fourth Affiliated Hospital of China Medical University, Shenyang, China
[7] Guangdong Provincial People's Hospital, Guangzhou, China

A Inspiration for $\mathcal{L}_{contra}$

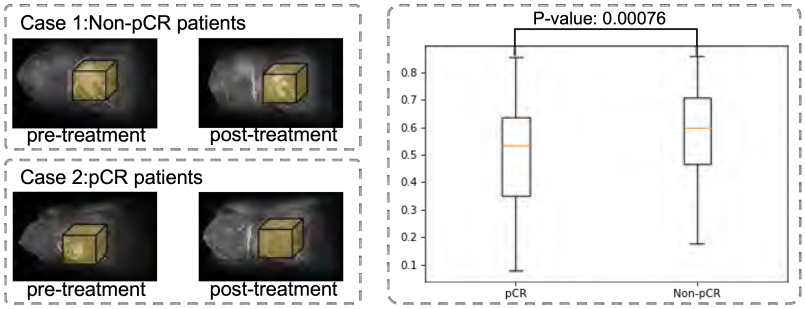

Fig. 1. The inspiration for multi-time MRIs contrastive learning. There are no invasive residual tumors visible on the post-treatment MRI scans of patients who achieve pCR. On the contrary, for patients who do not achieve pCR, tumors do not completely regress, and tiny tumors are still present on the MRI scans after treatment.The image features of pCR patients before and after treatment should ideally exhibit significant dissimilarity than that of Non-pCR patients.We randomly select 100 patients and calculate the similarity of MRIs before and after treatment, and the results showe significant differences between the pCR group and the non-pCR group.

B Data collection and preprocessing

Data collection In this study, 579 patients with breast cancer from our two collaborating hospitals were collected. The clinicopathological characteristics of the patients from the internal training set (n = 300), the internal validation set (n = 75), and the external test set (n = 204) are summarized in Table 1. MRI scanners' manufacturers of each cohort are provided (The in-house cohort A -Philips 1.5T Achieva and 3.0 T Ingenia, GE 1.5T Optima MR360, Siemens 3.0 T Verio; The in-house cohort B -GE 3.0T SIGNA Pioneer).

Table 1. Clinicopathologic Characteristics of Data Sets

Characteristics		Internal Training Set	Internal Validation Set	External Test Set
Age (years,mean±std)		48.52 ± 9.50	51.23 ± 9.78	50.17 ± 10.41
Pre-NAC clinical T stage				
	1	25	5	0
	2	184	49	5
	3	57	15	3
	4	32	6	2
	Unknown	2	0	194
Pre-NAC clinical N stage				
	0	106	21	0
	1	192	54	9
	Unknown	2	0	195
Molecular subtype				
	HER2+	89	18	67
	HR+/HER2-	167	44	101
	triple-negative	44	13	36
Treatment response				
	pCR	107	27	74
	non-pCR	193	48	130

Pathologic assessment Standard histopathologic analysis was conducted in each hospital for the pathologic assessment of response to NAC. pCR is defined as no residual invasive cancer in both the breast and axillary lymph nodes, while cancer in situ was allowed in some cases (ypT0/ypTis ypN0). Two site pathologists with more than 5 years of experience in breast pathology who were blinded to MRI findings were involved. If any disagreement, breast pathologists with at least a 10-year experience at each site reviewed it for interobserver confirmation.

Data preprocessing N4 bias field correction, resampling, histogram normalization and z-score are applied to MRI data. Tissue regions in WSI are automatically segmented by OTSU's method following Faisal et al.

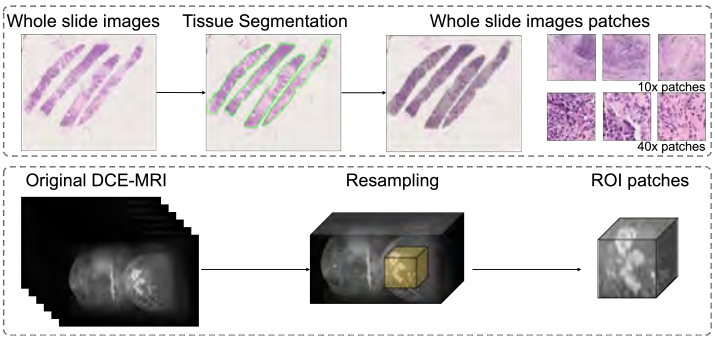

Fig. 2. Data preprocessing of pretreatment WSIs and MRIs.

C WSIs feature extraction

For WSIs feature extraction, we adopted an attention-based multiple-instance learning approach following Faisal et al. We use pre-trained vision transformers to extract patch features at different magnifications, denoted as $P^i = \{p^{i_j}\} \in \mathbb{R}^{i_J \times d}$, where i_J is the total number of patches for a patient. Finally, patch features are integrated by learned attention score a^{i_j} to obtain WSIs embeddings E^i for a patient.

$$\begin{aligned} a^{i_j} &= \frac{\exp\left\{W_{\mathrm{a},i}\left(\tanh\left(V_{\mathrm{a}}p^{i_j}\right) \odot \operatorname{sigm}\left(U_{\mathrm{a}}p^{i_j}\right)\right)\right\}}{\sum_{k=1}^{i_J}\exp\left\{W_{\mathrm{a},i}\left(\tanh\left(V_{\mathrm{a}}p^{i_k}\right) \odot \operatorname{sigm}\left(U_{\mathrm{a}}p^{i_k}\right)\right)\right\}} \\ E^i &= \sum_{j=1}^{i_J} a^{i_j} p^{i_j} \end{aligned} \tag{1}$$

D Evaluation

Compared to other multimodal fusion methods and single modality methods, M2Fusion yields improvement for pCR prediction with an AUC of 0.7346 in the internal validation set and 0.7992 in the external test set. Besides, FPVs are 0.2941 and 0.2759 respectively, lower than other methods. We have carefully check incorrect prediction cases. Some non-pCR cases are indeed identified as pCR. Breast cancer experts were consulted. They admitted that these are truly tough to identify. After NAC, tumors show regression and residual tiny tumors are hard to detect on MRI. Also, the surrounding tissue of the tumor may undergo significant changes after NAC, such as fibrosis, or necrosis, which may mask or mimic residual tumor tissue, making detection harder.

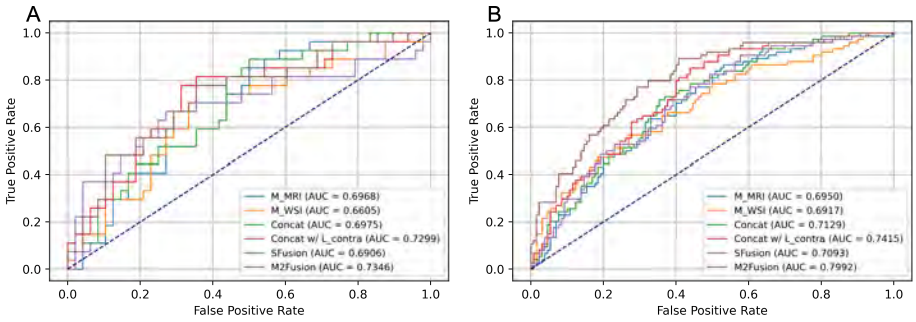

Fig. 3. Receiver operating characteristic (ROC) curves of different models **A.** ROC curves of different models in the internal validation set. **B.** ROC curves of different models in the external test set.