# Temporal Representation Learning of Phenotype Trajectories for pCR Prediction in Breast Cancer

Ivana Janíčková[1,2], Yen Y. Tan[3], Thomas H. Helbich[1], Konstantin Miloserdov[1,2,5], Zsuzsanna Bago-Horvath[4], Ulrike Heber[4], and Georg Langs[1,2,5]

[1] Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria
[2] Comprehensive Center for Artificial Intelligence in Medicine, Medical University of Vienna, Austria
[3] Department of Obstetrics and Gynecology, Medical University of Vienna, Austria
[4] Department of Pathology, Medical University of Vienna, Austria
[5] Christian Doppler Laboratory for Machine Learning Driven Precision Imaging, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria
ivana.janickova@meduniwien.ac.at, georg.langs@meduniwien.ac.at
https://www.cir.meduniwien.ac.at

**Abstract.** Effective therapy decisions require models that predict the individual response to treatment. This is challenging since the progression of disease and response to treatment vary substantially across patients. Here, we propose to learn a representation of the early dynamics of treatment response from imaging data to predict pathological complete response (pCR) in breast cancer patients undergoing neoadjuvant chemotherapy (NACT). The longitudinal change in magnetic resonance imaging (MRI) data of the breast forms trajectories in the latent space, serving as basis for prediction of successful response. The multi-task model represents appearance, fosters temporal continuity and accounts for the comparably high heterogeneity in the non-responder cohort.In experiments on the publicly available ISPY-2 dataset, a linear classifier in the latent trajectory space achieves a balanced accuracy of 0.761 using only pre-treatment data ($T_0$), 0.811 using early response ($T_0 + T_1$), and 0.861 using four imaging time points ($T_0 \rightarrow T_3$). The full code can be found here: https://github.com/cirmuw/temporal-representation-learning

**Keywords:** Temporal representation learning, Self-supervised learning, Breast Cancer

## 1 Introduction

Pathological complete response (pCR) to neoadjuvant chemotherapy (NACT) of breast cancer is a key marker of success determined on the basis of tissue resected during surgery [16]. Predicting pCR by assessing early response dynamics can steer treatment decisions. Even after concluded NACT, it may inform important choices such as forgoing surgery in case of expected pCR, given sufficient

prediction reliability. While single time point observations lack information on subtle, dynamic changes to treatment [10,19], longitudinal imaging may capture changes associated with individual treatment efficacy or disease progression.

Here, we propose a multi-task model to learn trajectory representations of imaging features observed during treatment. We show how a simple classifier can use this representation to predict future pCR with high accuracy. Our approach addresses the challenge of high inter-label similarity [13] and relatively substantial inter-individual variability not associated with treatment response. The dynamics of early response enables better prediction, while multi-task representation learning accounts for the response heterogeneity.

*Related work* Prior efforts to predict pCR in breast cancer imaging have used radiomics-based [9, 14, 16] and deep learning-based approaches [1, 3, 4, 10, 11, 19, 23]. These methods largely focus on single [1, 4, 11, 14] or two time point predictions [9, 19, 23], limiting their ability to capture the full temporal dynamics of tumor progression. Although some studies incorporate multiple time points for pCR prediction [3, 10], only [10] explicitly models temporal relationships using an LSTM layer [8]. Using three imaging time points, the model achieved AUC = 0.706, Sensitivity = 0.483 and Specificity = 0.773 [10]. Learning temporal relationships by creating individualized treatment progression trajectories or leveraging the learned temporal relationships to improve single time point predictions remains unexplored. The publicly available ISPY-2 dataset consists of series of MRIs taken before and during NACT [15, 18]. It provides an opportunity to address these limitations by modelling temporal dynamics more effectively. Existing methods typically classify patients as responders (positive) or non-responders (negative) [21–23]. For instance, [23] pre-trains a pCR prediction model by clustering pre- and post-NACT non-pCR images (assuming minimal change during NACT) while separating pre- and post-NACT pCR images (assuming greater change). However, this binary framework overlooks the heterogeneity among non-responders, including partial responders. The classification performance of this approach resulted in a test set AUC of 0.695 for the binary pCR label [23].

*Contribution* We propose a representation learning approach for image trajectories observed during treatment. We assume that, over time, responders change similarly, while non-responders are more heterogeneous since they include both partial responders and non-responders. The multi-task model learns to embed appearance change, while fostering temporal continuity by a dynamic-margin triplet loss adapted to nuanced temporal relationships. It aligns trajectories of responders to identify common temporal dynamics associated with successful treatment response, while accounting for the heterogeneity within the non-responder group. It avoids label-driven loss functions for non-responders and instead identifies response-specific patterns by aligning positive-outcome trajectories. A multi-task attention mechanism (MTAN) [17] enables the focus on feature changes associated with disease and treatment as opposed to comparably static inter individual variability of anatomy.

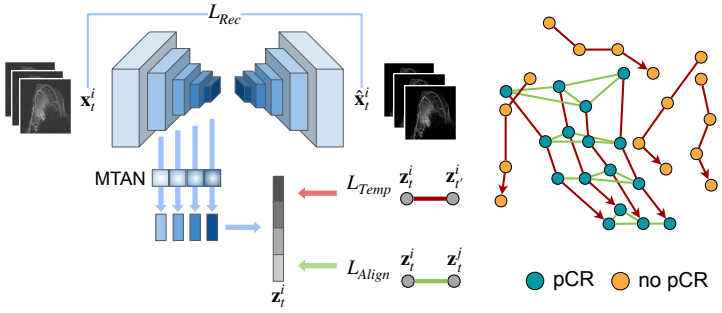


**Fig. 1. Method Overview** Multi-task representation learning balances reconstruction performance $L_{Rec}$ with temporal continuity of trajectories $L_{Temp}$, and alignment of changes in responders $L_{Align}$. A U-shaped denoising network extracts multi-scale features via its encoder. An MTAN module [17] steers attention across these tasks. The resulting trajectory representations are used for pCR prediction with a linear classifier.

Experiments on the multi-center ISPY-2 dataset [15, 18] demonstrate that the learned representation enables high prediction accuracy with a linear classifier, even with a single or only few time points. The MRI-based classification performance surpasses the state of the art as reported in [10, 23].

To our knowledge, this is the first pre-training method designed to (1) explicitly distinguish responder vs. non-responder progression patterns and (2) encode temporal relationships within ISPY-2. Unlike [10] our approach directly learns temporal dynamics for improved treatment progression modeling. Additionally, it leverages learned trajectories to refine single time point predictions, such as early pCR prediction at $T_0$, enhancing its ability to capture treatment response patterns.

## 2 Method

For each patient $i$, we observe a time-series of images $\mathbf{x}_t^i$ for $t = 0, \ldots, T$ acquired before ($t = 0$), during, and after treatment ($t = T$), and a label $y^i \in \{0, 1\}$, where 1 denotes a positive outcome (achieving pCR) and 0 a negative outcome (not achieving pCR). The multi-task training model for representing images integrates three loss functions: a triplet loss [20] with dynamic margin for temporal modeling $L_{Temp}$, a cosine similarity loss for responder alignment $L_{Align}$, and a reconstruction loss for robust image feature learning $L_{Rec}$.

A set of two distinct augmented examples is generated for each training example: $A_1(\mathbf{x}_t^i)$ and $A_2(\mathbf{x}_t^i)$. The final multiscale representations, derived from the encoder $f$, are then given by: $\mathbf{z}_t^i = f(A_1(\mathbf{x}_t^i))$, $\bar{\mathbf{z}}_t^i = f(A_2(\mathbf{x}_t^i))$ as shown in Fig. 1.

Our model is a U-shaped encoder-decoder network, where the multi-scale repre-

sentations are generated from the encoder. The representation tensors are normalized to a unit hypersphere and were used to predict the pCR in patients.

### 2.1 Constructing Multi-scale Visual Representations

We train a U-shaped encoder-decoder network to obtain a multi-scale representation of images, while at the same time fostering temporal continuity and alignment of trajectories observed in responders (Fig. 1). Fine-grained features are extracted from the encoder's multi-scale feature maps. These are then pooled, projected and concatenated into a multi-scale representation tensor $\mathbf{z}$. To ensure that the extracted features are anatomically relevant, we incorporate a reconstruction task:

$$L_{Rec} = \sum_{i \in N} \sum_{t \in T} E(\mathbf{x}_t^i, \hat{\mathbf{x}}_t^i) \tag{1}$$

Here, $\mathbf{x}_t^i$ is the target input example and $\hat{\mathbf{x}}_t^i$ is the denoised reconstruction generated from the noise-augmented input $A_1(\mathbf{x}_t^i)$. The loss function $E$ quantifies the reconstruction error using the mean squared error.

### 2.2 Learning Temporal Relationships

To adapt the triplet loss [20] for learning patient-level temporal relationships, we define an anchor-positive pair as representations of two different views at the same time point $t$: $a = \mathbf{z}_t^i$, $p = \bar{\mathbf{z}}_t^i$. The negative point is the image representation from the same patient at a different time point $n = \mathbf{z}_{t'}^i$ where $t' \neq t$. Instead of a fixed margin, we use a dynamically changing margin $m$, based on the relative difference between $t$ and $t'$. Additionally, we replace the standard distance metric with negative cosine similarity $d$. The final triplet loss is defined as follows, with $N$ denoting the total number of instances in a batch:

$$L_{Temp} = \sum_{i \in N} \sum_{t \in T} \sum_{t' \in T} max(d(\mathbf{z}_t^i, \bar{\mathbf{z}}_t^i) - d(\mathbf{z}_t^i, \mathbf{z}_{t'}^i) + m, 0) \tag{2}$$

### 2.3 Learning Shared Patterns of Change in Responders

In order to learn the shared patterns in responders' temporal trajectories, their representations are aligned by establishing correspondences in the latent space. The alignment is defined for two patients $(i, j)$ with $y^i = 1$ and $y^j = 1$ and their corresponding representation tensors $(\mathbf{z}_t^i, \mathbf{z}_t^j)$. The objective is to learn population-level response patterns by minimizing the distance between $\mathbf{z}_t^i$ and $\mathbf{z}_t^j$. The alignment loss is then defined for pairs representation as:

$$L_{Align} = \sum_{i \in N} \sum_{j \in N} \sum_{t \in T} \mathbb{I}(y_i = 1 \wedge y_j = 1) d(\mathbf{z}_t^i, \mathbf{z}_t^j)) \tag{3}$$

Here, $\mathbb{I}$ is an indicator function that ensures that the loss is only computed for pairs of positive examples. The expression $d(\mathbf{z}_t^i, \mathbf{z}_t^j)$ represents the negative cosine similarity between the two representations.

### 2.4 Overall Loss Function

The objective of the pre-training phase is to optimize the combined loss function:

$$L_{ART} = \begin{cases} L_{Align} + L_{Rec} + L_{Temp}, & \text{if } y = 1, \\ L_{Rec} + L_{Temp}, & \text{otherwise.} \end{cases} \tag{4}$$

The combined loss function ensures that the positioning of negative-outcome examples ($y = 0$) in the latent space is influenced only by the reconstruction and temporal components. In contrast, positive-outcome examples are further aligned at a population level through the supervision ($L_{Align}$, Fig. 1).

### 2.5 Feature Masking in Multi-Task Learning

We incorporate a learnable attention mask inspired by the MTAN module [17] to emphasize temporal changes and response-specific patterns in feature maps. It balances shared feature learning ($L_{Rec}$) with task-specific details ($L_{Align} + L_{Temp}$), refining representations to capture spatio-temporal changes during treatment.

## 3 Experiments and Results

**ISPY-2 Dataset** We used 585 patients from the public ISPY-2 dataset [15,18] with complete MRI scans at four NACT time points and all three DCE-derived maps: early enhancement ($PE_{\text{early}}$, 120–150 sec post-contrast), late enhancement ($PE_{\text{late}}$, ∼450 sec), and signal enhancement ratio ($SER = \frac{PE_{\text{early}}}{PE_{\text{late}}}$), enabling consistent longitudinal comparisons. These features capture contrast washout dynamics, offering insights into tumor biology and vascular properties [6]. Within this cohort, the proportion of patients achieving pCR is 33 %. To reduce memory usage, we generated axial-plane maximum intensity projections (MIPs) of the three DCE-derived volumes. The dataset was split into 70% training-, 10% validation-, and 20% test sets, stratified by pCR label. All volumes were resized to 256×256×256 and intensity-normalized to [0,1] before MIP generation.

**Implementation Details** The model is built on a UNet backbone [7] using MONAI's BasicUNet [2], initialized with features argument set to [16, 32, 64, 128, 256, 32]. The encoder-decoder structure was used for reconstruction (Fig. 1), while multi-scale encoder features were concatenated for temporal and response learning. We incorporated MTAN's masking strategy [17] to selectively refine relevant feature maps, ensuring better alignment with temporal and response dynamics. A two-layer MLP projector was then applied, resulting in a final feature size of 480. Pre-training was conducted using Adam optimizer [12] with a learning rate of 0.0001 and a batch size of 32 for 100 epochs. The triplet loss margin was dynamically set by encoding MRI time points in range of $[0, 1]$

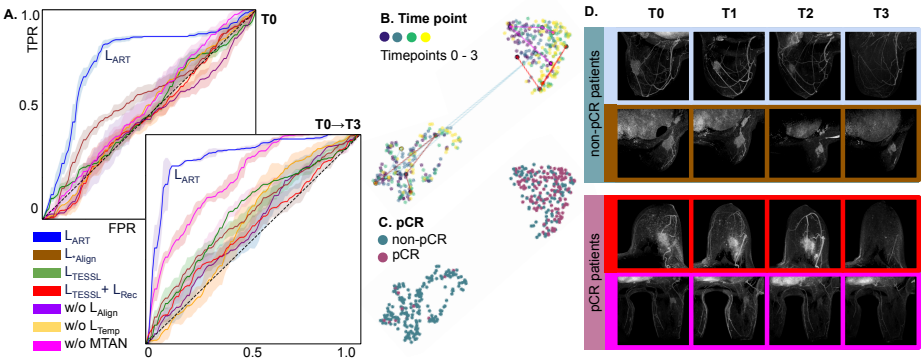


**Fig. 2. (A)** ROC for pCR predictions of different models. **(B)** UMAP projection of the test set data, colored by time point label, with plotted trajectories for two non-pCR patients and two pCR patients. **(C)** Same as B, colored by pCR label. **(D)** Image time-series for four patients, with visualized trajectories across all time points. Frame colors correspond to the trajectory colors.

with a step size of 0.25. For comparison, we used $\mathcal{L}_{TESSL}$, a time- and event-aware SSL strategy [21] introduced at MICCAI 2024, which, like our approach, incorporates both temporal and supervised signals during pre-training. Gradient accumulation was applied over 8 iterations with a batch size of 16 to simulate an effective batch size of 128, with Adam (learning rate = 0.15) as the optimizer. All models were pre-trained on full time-series data ($T_0 \rightarrow T_3$).

**Evaluation measures** For evaluation, we applied linear classifier to the frozen pre-trained features using `sklearn`'s LogisticRegression. Performance was assessed on the baseline time point ($T_0$) and the full time-series ($T_0 \rightarrow T_3$) over 10 runs, reporting the mean and standard deviation for the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (PRAUC) [5], and balanced accuracy. We further analyzed early response ($T_0 + T_1$) using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Model and parameter selection were performed on the validation set, results are reported for the test set.

### 3.1 Results

**Comparison with baseline method** We compared our approach with the state-of-the-art model [21] and two pre-training strategies: baseline $L_{TESSL}$, $L_{TESSL} + L_{Rec}$, and our proposed $L_{ART}$ (Eq. 4). Our method consistently outperformed the baseline across all metrics (Table 1, Fig. 2.A), achieving AUCROC of 0.892, PRAUC of 0.746, and Balanced Accuracy of 0.861 with the full time-series, and AUCROC of 0.764, PRAUC of 0.565, and Balanced Accuracy of 0.761 using only $T_0$ images. Bonferroni-corrected paired t-tests showed statisti-

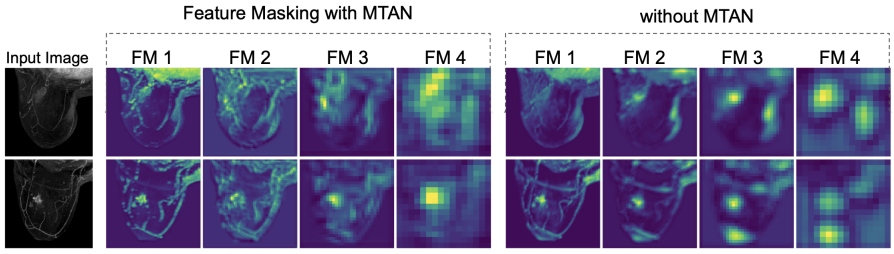


**Fig. 3.** Comparison of feature maps (FM) extracted from encoder trained with and without the MTAN module. Four feature maps visualize different levels of the encoder.

cally significant differences ($p<0.001$) for all metrics and evaluation time points ($T_0$, $T_0 \rightarrow T_3$).

**Ablation Study** We conducted an ablation study to assess the contributions of individual loss terms in our combined loss function $L_{\text{ART}}$ (Table 1, Fig. 2.A). Three key variations were examined: (1) removing the temporal loss (w/o $L_{\text{Temp}}$), (2) removing the alignment loss (w/o $L_{\text{Align}}$), and (3) applying the alignment loss across both pCR and non-pCR patients instead of exclusively to pCR cases ($L^*_{Align}$). Across all measures, all ablations resulted in a performance decrease. In addition, we evaluated the impact of the MTAN module [17]. As shown in Table 1, removing MTAN significantly reduced performance, confirming its role in enhancing feature informativeness for pCR classification. Fig. 3 illustrates how MTAN masking improves the attention focus on tumor regions. Lastly, we visualised the latent space of our model, highlighting temporal (Fig. 1.B) and outcome labels (Fig. 1.C).

**Early Response Prediction** Further analysis was performed to evaluate the prediction task for early response when only part of the time-series is available

**Table 1.** Linear evaluation results for the pCR prediction for early time point ($T_0$) and entire time-series ($T_0 \rightarrow T_3$) comparing a baseline approach $L_{TESSL}$ and $L_{TESSL}$ and $L_{Rec}$, ablated versions of the proposed approach (exluded components marked with w/o), and the proposed approach $L_{ART}$.

| Method | AUROC | | PRAUC | | Balanced Acc | |
|---|---|---|---|---|---|---|
| | $T_0$ | $T_0 \rightarrow T_3$ | $T_0$ | $T_0 \rightarrow T_3$ | $T_0$ | $T_0 \rightarrow T_3$ |
| $L_{TESSL}$ | 0.625±.01 | 0.556±.01 | 0.367±.01 | 0.449±.01 | 0.526±.01 | 0.565±.01 |
| $L_{TESSL} + L_{Rec}$ | 0.507±.02 | 0.556±.01 | 0.321±.01 | 0.413±.03 | 0.472±.02 | 0.556±.03 |
| w/o MTAN | 0.558±.02 | 0.783±.01 | 0.359±.01 | 0.613±.01 | 0.528±.03 | 0.697±.02 |
| w/o $L_{Temp}$ | 0.495±.02 | 0.562±.03 | 0.322±.02 | 0.346±.02 | 0.492±.03 | 0.518±.04 |
| w/o $L_{Align}$ | 0.575±.02 | 0.603±.01 | 0.388±.02 | 0.442±.02 | 0.593±.02 | 0.567±.02 |
| + $L^*_{Align}$ | 0.475±.02 | 0.554±.03 | 0.312±.01 | 0.356±.02 | 0.514±.02 | 0.528±.03 |
| $L_{ART}$ | **0.764±.01** | **0.892±.01** | **0.565±.02** | **0.746±.03** | **0.761±.01** | **0.861±.01** |

**Table 2.** Comparison of prediction accuracy for data at pre-treatment ($T_0$), early response $T_0 + T_1$, and full treatment timeline before surgery $T_0 \rightarrow T_3$.

| T | AUROC | PRAUC | Bal. Acc | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|
| $T_0$ | 0.764±.01 | 0.565±.02 | 0.761±.01 | 0.731±.02 | 0.762±.02 | 0.603±.02 | 0.852±.01 |
| $T_0 + T_1$ | 0.802±.01 | 0.649±.02 | 0.811±.02 | 0.769±.04 | 0.853±.01 | 0.721±.02 | 0.883±.02 |
| $T_0 \rightarrow T_3$ | **0.892±.01** | **0.746±.03** | **0.861±.01** | **0.846±.00** | **0.876±.02** | **0.772±.02** | **0.920±.01** |

($T_0$+$T_1$) as shown in Table 2. Adding $T_1$ improved performance across all metrics compared to pre-NACT prediction. Specificity and PPV reached values of 0.853 and 0.721, respectively, nearly matching those of the full time-series ($T_0 \rightarrow T_3$). This demonstrates the feasibility of predicting treatment outcomes from early response dynamics and the benefit of limited temporal information compared to static pre-treatment data ($T_0$).

## 4 Discussion

We propose a novel method that captures the temporal phenotypic dynamics of treatment response. It learns to represent response-specific patterns in serial MRI of BC patients undergoing NACT. The multi-task model generates individual temporal trajectories, aligning behavior in responders and representing image appearance using a joint loss function $L_{ART}$ balanced with an MTAN attention masking mechanism.

Comparative results underscore the contribution of the individual components of $L_{ART}$. Removing the temporal term $L_{Temp}$ leads to performance degradation, likely due to representational collapse. Omitting the responder alignment term $L_{Align}$ results in poor linear probing performance due to a lack of supervision during pre-training. At the same time, accounting for the heterogeneity of the non-responder group is crucial, as demonstrated by the drop in performance when aligning trajectories within both the responder- and non-responder groups using $L^*_{Align}$ as well as in the baseline $L_{TESSL}$. The role of temporal signal in pre-training is suggested by the decline in $T_0$ performance when excluding MTAN, in contrast to the full time-series performance (Fig. 2.A).

Linear classification results demonstrate that the learned representation carry relevant information for pCR prediction, outperforming previous methods [10,23] (see Sec. 1). Pre-training of representations using longitudinal data, also improves prediction using only single time point ($T_0$) and early response ($T_0 + T_1$) predictions, surpassing reported results.

Although ISPY-2 is a multi-center dataset, further validation on independent datasets would enhance the generalizability of our findings. Additionally, 3D CNNs would be ideal for volumetric information, but memory constraints and the size of the dataset limited us to 2D MIPs.

## 5 Conclusion

Predicting pCR in breast cancer patients is challenging due to the heterogeneity of individual response behavior. This study demonstrates that representing temporal dynamics can improve prediction accuracy. It identifies response-specific patterns in imaging data by balancing reconstruction, temporal continuity, and alignment of responder time-series. Evaluated with a frozen encoder and linear classifier, our method outperformed both the $L_{TESSL}$ loss for time-series [21] and prior results reported on the ISPY-2 dataset [10], highlighting the effectiveness of the pre-training approach. Predicting pCR using the full time-series has the potential to inform forgoing surgery. Results show that prediction based on early response is feasible, offering a perspective for early therapeutic adjustment.

**Acknowledgments.** This work was funded by the Vienna Science and Technology Fund (WWTF, PREDICTOME [10.47379/LS20065]) the European Union's Horizon Europe research and innovation programme under grant agreement No.101100633—EUCAIM, Austrian Federal Ministry of Labour and Economy, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association, and Siemens Healthineers.

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

## References

1. Bulut, G., Atilgan, H.I., Çınarer, G., Kılıç, K., Yıkar, D., Parlar, T.: Prediction of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer by using a deep learning model with 18F-FDG PET/CT. Plos one **18**(9), e0290543 (2023)
2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Comes, M.C., Fanizzi, A., Bove, S., Didonna, V., Diotaiuti, S., La Forgia, D., Latorre, A., Martinelli, E., Mencattini, A., Nardone, A., et al.: Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. Scientific Reports **11**(1), 14123 (2021)
4. Dammu, H., Ren, T., Duong, T.Q.: Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. Plos one **18**(1), e0280148 (2023)
5. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning (ICML). pp. 233–240. ACM (2006). https://doi.org/10.1145/1143844.1143874
6. El Khouli, R.H., Macura, K.J., Jacobs, M.A., Khalil, T.H., Kamel, I.R., Dwyer, A., Bluemke, D.A.: Dynamic contrast-enhanced MRI of the breast: quantitative method for kinetic curve type assessment. American Journal of Roentgenology **193**(4), W295–W300 (2009)
7. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al.: U-Net: deep learning for cell counting, detection, and morphometry. Nature methods **16**(1), 67–70 (2019)

8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
9. 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)
10. Jing, B., Wang, K., Schmitz, E., Tang, S., Li, Y., Zhang, Y., Wang, J.: Prediction of pathological complete response to chemotherapy for breast cancer using deep neural network with uncertainty quantification. Medical Physics **51**(12), 9385–9393 (2024)
11. Joo, S., Ko, E.S., Kwon, S., Jeon, E., Jung, H., Kim, J.Y., Chung, M.J., Im, Y.H.: Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Scientific reports **11**(1), 18800 (2021)
12. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Konz, N., Mazurowski, M.A.: The effect of intrinsic dataset properties on generalization: Unraveling learning differences between natural and medical images. arXiv preprint arXiv:2401.08865 (2024)
14. Li, P., Wang, X., Xu, C., Liu, C., Zheng, C., Fulham, M.J., Feng, D., Wang, L., Song, S., Huang, G.: 18 F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. European journal of nuclear medicine and molecular imaging **47**, 1116–1126 (2020)
15. Li, W., Newitt, D.C., Gibbs, J., Wilmes, L.J., Jones, E.F., Arasu, V.A., Strand, F., Onishi, N., Nguyen, A.A.T., Kornak, J., Joe, B.N., Price, E.R., Ojeda-Fournier, H., Eghtedari, M., Zamora, K.W., Woodard, S.A., Umphrey, H., Bernreuter, W., Nelson, M., ... Hylton, N.M.: I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2) (version 1) (2022). https://doi.org/10.7937/TCIA.D8Z0-9T85, https://doi.org/10.7937/TCIA.D8Z0-9T85, data set
16. Li, W., Newitt, D.C., Gibbs, J., Wilmes, L.J., Jones, E.F., Arasu, V.A., Strand, F., Onishi, N., Nguyen, A.A.T., Kornak, J., et al.: Predicting breast cancer response to neoadjuvant treatment using multi-feature MRI: results from the I-SPY 2 TRIAL. NPJ breast cancer **6**(1), 63 (2020)
17. Liu, S., Johns, E., Davison, A.J.: End-to-end multi-task learning with attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1871–1880 (2019)
18. Newitt, D.C., Partridge, S.C., Zhang, Z., Gibbs, J., Chenevert, T., Rosen, M., Bolan, P., Marques, H., Romanoff, J., Cimino, L., Joe, B.N., Umphrey, H., Ojeda-Fournier, H., Dogan, B., Oh, K.Y., Abe, H., Drukteinis, J., Esserman, L.J., Hylton, N.M.: ACRIN 6698/i-spy2 breast DWI [data set] (2021). https://doi.org/10.7937/TCIA.KK02-6D95, https://doi.org/10.7937/TCIA.KK02-6D95
19. Qu, Y.H., Zhu, H.T., Cao, K., Li, X.T., Ye, M., Sun, Y.S.: Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. Thoracic Cancer **11**(3), 651–658 (2020)
20. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)
21. Thrasher, J., Devkota, A., Tafti, A.P., Bhattarai, B., Gyawali, P., Initiative, A.D.N.: TE-SSL: Time and Event-Aware Self Supervised Learning for Alzheimer's Disease Progression Analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 324–333. Springer (2024)

22. Yue, H., Liu, J., Li, J., Kuang, H., Lang, J., Cheng, J., Peng, L., Han, Y., Bai, H., Wang, Y., et al.: MLDRL: Multi-loss disentangled representation learning for predicting esophageal cancer response to neoadjuvant chemoradiotherapy using longitudinal CT images. Medical image analysis **79**, 102423 (2022)
23. 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)